OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: May 3, 2003, 21:11:56 ; Search time 343.009 Seconds
(without alignments)
16426.693 Million cell updates/sec

Title: US-10-027-000-3
Perfect score: 2502
Sequence: 1 atgctgatattgatgttga..........attggtggtcggcgtgtaa 2502

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :   N_Geneseq_101002:*
     1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
     2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
     3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
     4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
     5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
     6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
     7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
     8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
     9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
    10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
    11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
    12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
    13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
    14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
    15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
    16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
    17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
    18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
    19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
    20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
    21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
    22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
    23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
    24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description
1	260.8	10.4	588	21	AAF15000	Trichoderma reesei
2	184.4	7.4	3241	24	AAS18442	Contig 93 DNA enco
3	179	7.2	2401	21	AAA75634	Nucleotide sequenc
4	179	7.2	2401	21	AAZ56002	Nucleotide sequenc
5	177.4	7.1	1145	16	AAT04785	Contig 001 from co
6	175.2	7.0	2430	21	AAZ27294	Cellobiase gene fr
7	175.2	7.0	2430	24	AAD39052	S. venezuelae macr
8	175.2	7.0	12441	21	AAZ87284	S. venezuelae venez
9	175.2	7.0	13613	24	AAD39043	Streptomyces venez
10	167.2	6.7	13613	21	AA287319	S. venezuelae deso
11	161	6.4	2256	17	AAT32999	Chimaeric thermost
12	138.6	5.5	50937	21	AAA09469	Streptococcus olea
13	135.8	5.4	2166	18	AAT93682	Thermotoga maritim
14	135.8	5.4	2166	19	AAV36911	Thermotoga maritim
15	111.6	4.5	4403765	22	AA199683	Mycobacterium tube
16	111.6	4.5	4411529	22	AA199682	Mycobacterium tube
17	103	4.1	985	24	ABQ44816	Oligonucleotide fo
18	103	4.1	985	24	ABQ44817	Oligonucleotide fo
19	101.4	4.1	16836	19	ABQ04837	Acetobacter xylinu
20	98.6	3.9	3849	22	AAN80309	S. chrysomallus ac
21	93.4	3.7	2271	11	AAQ06844	Amylase gene from
22	91.2	3.6	2291	9	AAF25795	Entire amylase gen
23	85.8	3.4	12588	15	AAV52831	Sequence encoding
24	85.4	3.4	65140	22	AAQ63293	Streptomyces nours
25	85.4	3.4	125401	22	AAO56859	Streptomyces nours
26	84.4	3.4	390	14	AAQ21833	Streptomyces sp.
27	84.4	3.4	390	13	AAD17184	Randomising oligon
28	84.4	3.4	390	24	AAD17186	PCR primer for 5'
29	84.4	3.4	390	22	AAF76910	Sequence containin
30	83.8	3.3	3032	21	AAI72775	Oligo #7 for cloni
31	83.8	3.3	3033	21	AAQ25325	Sequence encoding
32	83.8	3.3	3033	21	AAQ63953	Trichoderma reesei
33	83.8	3.2	1140	19	AAZ38879	DNA encoding a bet
34	80.4	3.2	7498	24	AAV41733	Trichoderma reesei
35	80.4	3.2	2064	14	AAH21757	Trichoderma reesei
36	80.2	3.2	1661	22	ABK91527	Codon-optimised RA
37	80.2	3.1	3269	16	AAT06024	Modified HIV prote
38	78.6	3.1	2064	16	AAQ52638	Corynebacterium gl
39	78.6	3.1	1294	20	AA206824	Streptomyces fradi
40	78.2	3.1	1294	16	AAH74537	cDNA encoding aven
41	78.2	3.1	1521	17	AAO85043	Streptomyces albid
42	78.2	3.1	1521	19	AAV11824	Nucleotide sequenc
43	78.2	3.1	1521	16	AAT62676	3-Hydroxysteroid-o
44	78.2	3.1	1647	17	AAT62675	3-hydroxysteroid-o
45	78.2	3.1	1647	19	AAV11823	Streptomyces sp. s

ALIGNMENTS

RESULT 1
AAF15000

AAF15000 standard; cDNA; 588 BP.
XX
ID   AAF15000 standard; cDNA; 588 BP.
XX
AC   AAF15000;
XX
DT   13-MAR-2001 (first entry)
XX
DE   Trichoderma reesei EST SEQ ID NO:7523.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
OS   Trichoderma reesei.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999; 99US-0273623.
XX
PA   (NOVO) NOVO NORDISK BIOTECH INC.
PA   (NOVO) NOVO NORDISK AS.
XX
PI   Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR   WPI; 2000-594572/56.

Left column

```
XX
PS   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
     Claim 89; Page 3038-3039; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 588 BP; 130 A; 167 C; 157 G; 109 T; 25 other;

Query Match             10.4%;  Score 260.8;  DB 21;  Length 588;
Best Local Similarity   92.8%;
Matches   283;  Conservative    0;  Mismatches   20;  Indels    2;  Gaps    1;

Qy   1870 ATCCAGGCCCTGGTACGGCCGGCAACGAGACGAGGCCAACTCCATTGCCCGACGTCGTCTTTGGC 1929
          ||||||||||||| ||||||| ||| ||||||||| ||||||||||||||||||||||||||||||
Db      1 ATCCAGGCCCTGGGTACGGCGGGGNAACGAGACGCGGCCAACTCCATTGCCCGACGTCGTCTTTGGC 60

Qy   1930 GACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACACCCC 1989
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     61 GACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACACACCCC 120

Qy   1990 GCGGTTCTCAACTTCCGCACCGAGGCGCCGGCGACCGACGTCTACGTC 2049
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCGGTTCTCAACTTCCGCACCGAGGCGCCGGCGACCGACGTCTACGTC 180

Qy   2050 GGGTACAGGTACTACGAGGTTCCCGACAAGGACGTCAAATTTCCCCTTTGGCCACGGCCTG 2109
          |||||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Db    181 GGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAAATTTCCCCTGTACGGGCACGGGCTG 240

Qy   2110 TCCTTACACCACCACTTTT--GCCTTTTCCAATCTCTCCCGGTGTCTCACAAGGACGGCAAGCTGA 2167
          |||| |||||||||||||||  |||||||||||||||||| ||||||||||||||||||||||||
Db    241 TCCTACACCACCACTTTTTTGCCTTTTCCAATCTCTTCCCGGGGTCTTAACAANGACGACGGNAAA 300

Qy   2168 GCGGTG 2172
          |||||
Db    301 GCTTG 305


RESULT 2
AAS18442
ID   AAS18442  standard; DNA; 3241 BP.
XX
AC   AAS18442;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Contig 93 DNA encoding S. narbonensis polyketide synthase.
XX
XX
```

(handwritten annotation: "Cannot we tell due to In re Duvall" with signature)

Right column

```
KW   Narbonolide polyketide synthase; PKS; narbomycin modification enzyme;
KW   erythromycin; rapamycin; tylosin; picromycin; methylmycin; neomethymycin;
KW   agriculture; ds.
XX
OS   Streptomyces narbonensis.
XX
PN   US6303767-B1.
XX
PD   16-OCT-2001.
XX
PF   05-NOV-1999;   99US-0434288.
XX
PR   05-NOV-1998;   98US-1070093P.
PR   27-MAY-1999;   99US-0320878.
XX
PA   (KOSA-)  KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  McDaniel R;
XX
DR   WPI; 2002-055495/09.
XX
     Claim 1; Column 20-22; 24pp; English.
XX
CC   The present invention relates to recombinant DNA vectors (cosmids)
CC   that encode for the narbonolide polyketide synthase (PKS) enzyme and
CC   various narbomycin modification enzymes from Streptomyces narbonensis.
CC   The recombinant DNA vectors can be used to produce recombinant ketide
CC   synthases and a variety of different polyketides (e.g. erythromycin,
CC   rapamycin, tylosin, narbomycin, picromycin, methylmycin and
CC   neomethymycin) for use in agriculture, medicine and health. The
CC   recombinant vectors may be used to produce polyketides in relatively
CC   high yields. AAS18432-AAS18443 represent contig DNA sequences that
CC   encode for S. narbonensis PKS enzymes.
XX
SQ   Sequence 3241 BP; 520 A; 1174 C; 1125 G; 422 T; 0 other;

Query Match             7.4%;   Score 184.4;  DB 24;  Length 3241;
Best Local Similarity   48.2%;
Matches  1028;  Conservative    0;  Mismatches  956;  Indels  150;  Gaps   12;

Qy    206 CCCTCGGTTCACACATTCAACCANAACTTCGCTCGAAGAGGCAGGTAAGATGANGCGGCAAAG 265
          ||| || | || |||  |  || |  |  |||  |  ||  |||||  ||  || |||| |||
Db   1151 CCCTGGCCCAGCACCTTCGACGAGACTCCATGGCCGCACAGCTACGGCACGGGTCATGGCCCGCG 1210

Qy    266 AGGCCATCGCTAAGAGTGCGCAINGTGATCCTCGGCCGCACTATCCACATGCGAACGCTCCC 325
          | ||||| || | | |    |  || || || |||||||| |  || ||||  ||   |  |
Db   1211 ACGGACCGCGCCGCTGGGCCCCAGGACACATGGTTCTCGGGCCCGATGATGAACAACATCCGGGTGC 1270

Qy    326 CTCTCGGTGGACGTGGCGTTCGAGTCGATTGGTGAGGATCCGGTTCCTGGCGGGGCTTGGGAG 385
          |  | || |||   | ||   |  |||   |||||| | | ||||| ||    | || |||
Db   1271 CACACGGCGGCGCGGACTACGAACCTTCACGGAGCCCCTGGTCTCCTCCGCGCACCG 1330

Qy    386 CTGCGGGCTCTCATCCGCGGCATTCAGAGCACTGAAGGTGCAGGCTACGGATCAAGCACTTTT 445
          | || || |||||| ||  ||||||| || | | | | | ||| |||||| |||||||| |
Db   1331 CGGTCGCCCCAGATCAAGGGCATCCACGGGTGCGGGGTCTGATGACCACGGCCAAGCACTTCG 1390

Qy    446 TGTGCAANGATCAGGAGGACAGGCCGCATGGATGGTGCAGAGACATCGTCACGGAGCGGGCTC 505
          |  |||| || | ||   |||| || ||   |  |||| | | | |||   || | || |
Db   1391 CGGCCACAACGACAGGAGAAACAACCGCTTCAGCGGT---GCGTCCTCGAAGGCCGGCGGCTCCT 1450

Qy    506 TCCGGTGAAATCTACGGCACTCCCGTTCCAGATTGCGCTGCGGAGACTCCCAGCCCGGTGCGT 565
          |||| |||| | | | | | ||||| ||||||  ||   |||||  | ||   ||| ||| |
Db   1451 TCCGCGAGATCAGTTCCCGGGGTTCAG---GCGTCCTCGAAGGCCGGCGGCTCCT 1507

Qy    566 TCATGACGGCGGTACAAATGGCATCAAATGGCGTGTTGTCGCAGCGGAGAAACCCTAAATATCTTG 625
          | | ||| || |||   || ||||| |||  ||    | | |||||  || | | || |
Db   1508 TCATGTGTGCCTATATACGGCGTCAAACGGGCAAGCCGTCCGTCCTGCGGCAACGACGAGCGTGCTCA 1567

Qy    626 ATGGGATGCTTCGAAAGGAATGGGTTGGAGGGCCCTPAATCATGACGGACTGGTACGGCA 685
          | |||||||  |  ||  |||| |||  ||| |||||||||| | | | | || ||| |
Db          ...
```

The page consists of a lengthy DNA sequence alignment (Qy/Db paired lines with match bars and flanking coordinate numbers) followed by a database record block.

XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT .
XX
PS Disclosure; Columns 39-40; 117pp; English.
XX
CC The present sequence is used to produce the recombinant DNA compounds
CC of the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 2401 BP; 405 A; 859 C; 828 G; 308 T; 1 other;

Query Match   7.28;   Score 179; DB 21; Length 2401;
Best Local Similarity   47.78;   Pred. No. 5.6e-24;
Matches 1018; Conservative   0; Mismatches   966; Indels 150; Gaps 11;

```
Db     776 --------GGGGTCGAGCTCCCCGGCGACGTCCCCGAAAGGCGAGCCCTCGCCGCGGCCA 827
Qy     206 CGCTCCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGCCAAAG 265
            | ||  | |||||   |||    |||  || | |    ||| ||||||   | ||   |||
Db     255 CCCTGGCCAGCACCTTGACGACACCATGGCCGACAGCTACGGCAAGGTCAATGGGCGCGCG 314
Qy     266 AGGCCAATCGCTAAGAGTGCCGATCTCGGCCCGACTATCACAATGCAAACGCTCCC 325
            ||| |||  |||| | | || | |||||||  ||| || | || |||||  || ||
Db     315 ACGGTCCGCGCCTCAACCAGGACAGTGGTCCTGGGCCCGATGAATGAACACATCCGGGTGC 374
Qy     326 CTCTCGGTGGACGTGGCCTTCGAGTCGATTGGTGAGGATCCGTTCCTCGGCGGGGCTTGGGAG 385
            | |||   |||| |||||||| |||| |||  ||||| ||||||  | || |||  |||
Db     375 CGCACGGCCGCGGAACTACGAGACCTTCAGCAGGACCCCCTGGTTCTCCTCCGCCACCG 434
Qy     386 CTGCGGCCTCTCATCCCCGGCATTCAGAGCACTGGAGTGCAGGCTACGATTCAAGCACTTTT 445
            ||  ||||| ||| |  |||   |||| || | ||||||  |||| ||| |||| || ||
Db     435 CGGTCGCCCAGATCAAGGCATCCAGGGTGCCGGGTTCTTGATGACCACGGCCACTTCG 494
Qy     446 TGTGCAATGATGATCAGTAGGAGGACAGGCGCATGAATGGTGCAGAGAATCGTCACGGAGCCGGGGCTC 505
            | |||  |||  ||||||| | | | || || |||  || ||||  | || ||| ||||
Db     495 CGGCCAACAACCAGGAGAACAACCGCTTCTCCGGTGAACGCCAATGTCAGCAGACGAGCACGC 554
Qy     506 TCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCCAGCCGGGGTCGT 565
            | | || |||  ||| ||||  | || |||||| || ||   ||| |||||  ||||
Db     555 TCCGCGAGATCGAGTTCCCGGGCGTTCGAG---GCGTCCTCCAAAGGCCGCGGGACCTCCT 611
Qy     566 TCATGACGGCGGTACAAATGGGCATCAAATGGCGTGTCGTGCACGCGAGAACCCTAAATATCTTG 625
            ||  || ||| |||   ||||||   ||   |||  ||| |||| ||||||   || ||||
Db     612 TCATGTGTGCCTACAACGGGCTCAACGAGCCGTCTGCGGCACGACCAGAGCTCCTCA 671
Qy     626 ATGGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCGACTGGGTACGGCCA 685
            ||| |||||| | || ||| |||| || |||||| || ||   ||||||  |
Db     672 ACAACGTGCTGCGCCACGCCAGTGGGGGCCTTCCAGGCTGGGTGAGTGTCCGACTGGCTCGGCCA 731
Qy     686 CATACAGTACCACAGAGAAGCCGTTGTTGCGCAGGCCGCTCGACCTCGACCTCACTCCAC 745
            | || | | | ||||    ||| ||||| |||| ||||||||||| || ||||  |||
Db     732 C---CCCGGGCACCGACCCCATCACCAAAGGGCCCTCGACCAGGAGATG------------ 775
Qy     746 GCTTCCGAGGAGGAGAAACACTCAAGTTCAACGGAAAGCCCCTTTATCCACGTCA 805
            |||    | |||||||||    ||  || || |  | ||  |||   ||||| ||
Db     806 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCCTGGCCTCG 865
```

PT   New recombinant DNA encoding a domain of narbonolide polyketide
PT   synthase, for production of ketolide antibiotics -
XX
PS   Disclosure; Page 34-35; 98pp; English.
XX
CC   This is contig 001 from the recombinant cosmid pKOS023-27 DNA sequence
CC   (see AAZ56001) which contains a Streptomyces venezuelae DNA insert. The
CC   cosmid contains open reading frames which encode the various modules of
CC   the narbonolide polyketide synthase (PKS). The invention relates to
CC   recombinant DNA containing a coding sequence for a narbonolide PKS.
CC   Polyketides are compounds synthesised from 2-carbon units through a
CC   series of condensations and subsequent modifications. Modular PKSs are
CC   responsible for the production of many antibiotics including picromycin.
CC   The narbonolide PKS consists of a loading module, six extender modules,
CC   and two thioester domains. Four proteins make up the narbonolide PKS
CC   (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC   and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC   PICAIII includes extender module 5 and PICAIV includes extender module 6
CC   and a type II thioester domain. The second type II thioesterase
CC   domain is found on the PICB protein. The nucleotide sequences encoding
CC   all of these proteins can be isolated in recombinant form from the
CC   recombinant cosmid pKOS023-27. Narbonolide is desosaminylated in S.
CC   venezuelae to yield narbomycin, the desosaminyl transferase enzyme is
CC   required for this conversion, and the desosamine biosynthetic genes are
CC   also found in cosmid pKOS023-27. The recombinant DNA of the invention is
CC   used to express, in transformed cells, narbonolide (or its derivatives)
CC   or other ketolides (particularly hybrids), which may then be converted
CC   (e.g. by other enzymes (particularly expressed in the same hosts) to
CC   polyketide antibiotics or their intermediates. The antibiotics are useful
CC   in human or veterinary medicine.
XX
SQ   Sequence 2401 BP; 405 A; 859 C; 828 G; 308 T; 1 other;

Query Match   7.2%; Score 179; DB 21; Length 2401;
Best Local Similarity   47.7%;
Matches 1018; Conservative   0; Mismatches 966; Indels 150; Gaps 11;

Qy   206 CGCTCGGTTCACATTCAACCAAAACTCGCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 265
          ||||| |||| |||| ||||| ||| ||| | || || || |||||| |||| |||
Db   255 CCCTGGCCAGCACCTTCGACGACACCATGGGCGCAGAAGCTACGGCAAGGTCATGGGCGCGG 314

Qy   266 AGGCCATGCGTAAGGAGTGCCCATGTGAATCCTCGGCCCCGGACTATCAACATGCAACGCTCCC 325
          |||| |||| || ||| || || | || || | || || || |||| ||
Db   315 ACGGTCGCGCGCCGTCAACCAGGACACGACGTCCCTGGGCGCCCGATGAATGAACAACATCCGGTGC 374

Qy   326 CTCTCGGTGGACGTGGCCTTCGAGTCCGATTGGTGAGGATCCGTTCGTGGCCGGCGGC**TGGGAG 385
          ||||| || || | || || ||| | | || || |||| || | || ||||
Db   375 CGCACGGCGCGGCGGAACTACGAGACCTTCAGGCAGAGCCCCCTGGTTCTCTCGCGCCACCG 434

Qy   386 CTGCGACGGCTCTCATCCCGGCAATTCAGAGCACTGGGACTGCAGGCTACGATCAAGCACTTTT 445
          || | ||||| || | | | || ||| || || || |||| | || | |||||
Db   435 CGGTCGCCCAGATCAAGGGCATCCAGGGTGCGGGCTCTGATGACCACGGCCAAGCACTTCG 494

Qy   446 TGTGCAATGATCAGGAGGAGGCACAGGCCGCCATGAAGGTGCAGAGACCATCGTCACGGGCGGGGGCTC 505
          ||| | |||| || || || || || || |||||| ||| ||| | || |||| ||
Db   495 CGGCCAACAACCAGGAGAACAACCGCTTCTCCGTGAACGCCAAATGTCGACCAGCACAACGC 554

Qy   506 TCCGTGAAAATTACGGCACTCCCGTTCCAGATTGCTGTCGCGAGACTCCCACCCGGGTGCGT 565
          ||||||| | || || |||| ||||| || | || || ||||| || ||
Db   555 TCCGCGAGATCGAGTTCCCCGGCGTTCGAG---GCGTCCTCCAAGGCGGCGGCGCGGGCTCCT 611

Qy   566 TCATGACGGCGGTACAATGGCATCAATGGCCGGTGTCGTGCAGCGGAGAACCCTAAAATATCTTG 625
          |||| |||||||| || |||| ||| ||| ||| || |||| |||| ||| || |||
Db   612 TCATGTGTGCCTACAACGCGGGCTCAACGGGAAGCCGTGCTTGCGGCAACGACGAGCTCCTCA 671

Qy   626 ATGGGAATGCTTCGAAAAGGAATATGGGGTTGGGATGGCCTAATCATGAGCGACGACTGGGTACGGCA 685
          |||| ||| || | | ||| || ||| | || | ||||||||||| ||||| |
Db   672 ACAACCTGCTCGCGCACGCAGTGGGCGCTTCCAGGCGTGGGTGCATGTCCGACTGGCCTCGCCA 731

Qy   686 CATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGCCCGGACCTTCCAC 745
          ||| | | ||| || || || || || ||||||||||| || |||||
Db   732 C---CCCGGGCCACCGACACGGCCATCACCAAGGGCCTCGACCAGGAGATG--------- 775

RESULT 4

AA256002
ID   AAZ56002 standard; DNA; 2401 BP.
XX
AC   AAZ56002;
XX
DT   23-MAR-2000 (first entry)
XX
DE   Contig 001 from cosmid pKOS023-27 from Streptomyces venezuelae.
XX
KW   Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; contig 001;
KW   ketolide; 1 beta glucosidase; antibiotic production; narbomycin;
KW   picromycin; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key   Location/Qualifiers
FT   CDS   80..2389
FT   /*tag= a
FT   /product= 1_beta_glucosidase

PN   WO961599-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999; 99WO-US11814.
XX
PR   28-MAY-1998; 98US-0087080.
PR   28-AUG-1998; 98US-0141908.
PR   22-SEP-1998; 98US-0100880.
PR   08-FEB-1999; 99US-0119139.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR   WPI; 2000-072618/06.
DR   P-PSDB; AAY67212.
XX

Qy   1886 GCGGCAACGAGACGGGCAAACTCATTCCGACGTACGTCTTTGGGACTACAACCCTCGG 1945
          ||||||||||||||| | | || || || ||||| |||| ||
Db   1797 CGGGCAACGAGACGGGCCGAGGCCACCGGCGTGCTCTACGGTGACGTCAAACCCAAGCG 1856

Qy   1946 GCAAGCTGTCCCTCAGCTTCCC------------AAGGCCTGCAGGACAACCCCGGTTTC 1996
          ||||||| ||| | || |||| || |||| |||| || ||
Db   1857 GCAAGCTCACCGACAAGCTTCCCGGCGCCGAGAACACGGTGCCGGCGACCCGA 1916

Qy   1997 TCAACTTCCGACACGAGCTCCGCCGGACGGCTGCGGGGGTCGGGGTACA 2056
          | ||| |||| ||||||| | |||| || || | ||||||
Db   1917 CCAGCTACCCGGGCGTCGACAACAGCAGCGTACCGACGGCATCACCGTCGGGGTACC 1976

Qy   2057 GGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCACGGCCGGTCCTGTCCTACA 2116
          | || ||| ||||| |||| |||| || | |||| ||| ||||||| ||||||||
Db   1977 GCTGGTTCGACAAGGAGGAAACGTCAAGCCGCCGTTCCGTTGGGCACGGCCTGTGCTGTACA 2036

Qy   2117 CCACTTTTGCCTTTTCCAAATCTCTCCGTGTGTCTCACAAGGAC---GGCAAGCTGAGCGTGT 2173
          || |||| || ||| ||||| ||||||| ||| ||| || |||| |||||||| |
Db   2037 CCTCGTTCACGGAGGAGCCCGACCTCGTTCGTACGTCACGGGTGGTCTGAAGGTCA 2096

Qy   2174 CCCTCTCCGTGAAGAACACCGGCTTCGGTTCCGGCCACGTGGGCCCAGCTCTACGTCA 2233
          ||||||||||| ||| | ||||| || |||||| ||| || |||| |||||
Db   2097 CGGTCACGGTCCGCACGGGGGAAGCGGCGGTCGTTCCAGGCGGTACCTCG 2156

Qy   2234 AGCCCCTTCCAACCGCCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGTTCGCAAAGG 2293
          || ||| || || ||||| ||| || || | |||| || || || ||||
Db   2157 GTGCCAGCCCGAACGTAACGGCTCCCGCAGGGCGAAGAGAAGCTCGTGGGCTTCACGAAGG 2216

Qy   2294 TCGAACTGCAGCCCCGGCGGAGAGCAGAGAAGGCGGTGAC 2327
          || |||||||||| || |||| ||||||||||||||||||
Db   2217 TCTCGGTTCCGCGCCGCGGGGCGAGGCGAAGACGGTGAC 2250

```
Qy    746 GCTTCCGAGGAGAAACACTCAAGTTCAACGTTCCAACGCTCA 805
              |||           |||||||||||||||||||| ||
Db    776 --------GGCGTCGAGCTTCCCCGCGACGTCCCGCCGCCCA 827

Qy    806 TTGACCAAGAGGGCTAGGGAAGTTCTTCGTCAAGAGAGTTGA 865
              ||||       ||||   ||||||||||||     |||||
Db    828 AGTTCTTCGGCGA---GGCGCTGAAGACGGGCCGTCCTGAAACGGCACGGCGCGG 884

Qy    866 CGGAGAACGGCCCCGAGAGACGACTGTCAACAACACCCCGAAAACGGCAGCTCTTCCTCCCGGA 925
              ||||||||||        ||   ||  ||||        |||||||||| |||||||||
Db    885 CCGTGACGCGGTCGGCGACGGGGATCGTCGGCCAGATGGGAAGTTCGGTCGGTCTGCTTCTCG 944

Qy    926 AGGTTGGCAACGAGGGCAATCGTGCTGCTGAAGAAACGAACAACGTTCTGCCCTTGACCA 985
              ||||   |||||||       ||||||||    ||       ||||||||||||
Db    945 CCACTCCGGCGCCGCGGCCCGGCGGGACGGGGTGCCCAGCGCGGGTGTCCCGCGAAGG 1004

Qy    986 AGAAGAAGAAGACGGCTGATTGTCGGGCCCCAAGCCACAGCCACACGGCGGAG 1045
              ||||||||||||||||        ||||||||        |||   |||||||
Db   1005 TCGGCGAGAAACGGCGGGGTGCTCCTCGGGCCAACGAGGGCCCTGCGCGGCGTG 1064

Qy   1046 GCTCTGGCCGCCACTCAAGGGCCTACTACGGCAGTCACTCCCTTTGACGGGCCTCAGCAAGCAGC 1105
              ||||||||       ||||||||   ||   |||||||   |||||||||||   ||||||
Db   1065 ACGGCGGAAGAGCATCGGGTCATCGGGCCGGCGTCGAC-CCCAAGGTCACCGGC 1123

Qy   1106 TCGAGACGGCGCCATCGTACACGCGTGGCGGCCTACACCAGTTCCTCCCATTCTAGGCG 1165
              |||||       |||   ||||||||||||        |||||||||   |||||||
Db   1124 CTGGGCGGCAGCGCCCACGACTCGGCGGGCCACTCGACGACCATCAAGGCC 1183

Qy   1166 AGCAGTTGCTTCACGCCCGACGGGCGCTCGCGGGAGGGTCTTCACGAGCCC 1225
              |||||||||||||||||||||||||||||        ||| |||||||
Db   1184 CGCCGCGGGTGCGGGTGCGACGGTGCGACGAGACGGGGACCTTTGGGGACGCAG 1243

Qy   1226 CTGGTACCCCTAACCGCCCAGCACATTGACGAGCTCTTCTTCACCAAGACGGACATGCACC 1285
              ||||||||       ||||||||||||||||        |||   |||
Db   1244 ATCCCGGCGGGGAACCTCAGC----------CCGGCGTTCAACC 1277

Qy   1286 TGGTTGGGACTACTACCACCCCAAGGCGGGCGAGACACGTGGTTACGCCGACATGGAGGGCACGT 1345
              |||||||   |||   |||||||||||||||||||||   |||||||||||||||||||||||
Db   1278 AGGGCCACCAGCTCGACGGCCGAAGGCGGCGTTGACGACGGCACGGCGGCTGACCCAGGCTGC 1337

Qy   1346 ACACCGCCGACGAGGACGCTCACCTTACGAGCTCGGCCCTCGTCGTTGCGGCACGGCAAAGG 1405
              ||||||       |||||||||||   ||||        ||||||||   ||||||||||
Db   1338 CCGGCGACGGCGGAGTACCCCATCACCACTACACCTCAAAGGCGACAACCATCG 1396

Qy   1406 CGTACGTAGGAGCGACCAGCTCGTCGTCGACAACGGCCACCACAAGCAGGTCCCCGCGATGCCT 1465
              |||       ||||||||||   |||||        ||||||||   |||||||||||
Db   1397 --------------------CTCGGCAGCCACA 1409

Qy   1466 TCTTCGGCGTCCGACCACCCCGCGGGAGGAAGACAGGGCCCGCCATCAATCTCGTCAAAGGGCAACACGT 1525
              ||||||       ||||||||||||||||||||||||||||||||   |||||||||||||||
Db   1410 CCATCGAGGCCGGTGGCTCTACGGCAGCCCGCGGCTTCAAAGCTGACCAAGG 1469

Qy   1526 ACACGCCGACGAGGAGCACTCACCTTACGGAGCTCCGCCTCGTCGTGCGGCACCGCAAAGG 1585
              ||||||||||       |||||||       ||||||||||||||||||||   ||||||
Db   1470 GCACGCACAAGCT-------------------CACGATCTCGGGGCTTGCGGATGA 1505

Qy   1586 TCCCCGGCACGCGCTCCCTCCGGGTGGCTTGCAAGGTCATTGACGACCAGGCCGAAA 1645
              |||||||   ||||||||||   |||||||||||   ||||||||||||||||||||
Db   1506 GTGCCACCCGCGTCTCCCTGGACGTGGGCTGGGTNACGCGGGCGACGGCGACGA 1565

Qy   1646 TCGAAAAGTCGGTGCCCCTTCGCCAAGGAGCACGACCAGGTCATCATCTGCGGGGGCCTTA 1705
              ||||||||       |||||||||||||||||   |||||||||||   |||||||
Db   1566 TCGGGAAGCCGGTGGAGTCGGGCGGGAAGGCCGTACGGCGGTCGTCTTCGC------- 1617

Qy   1706 ACGGCGACTGGGAGACGAGGGCGGGCGACGACGCGGCAAGTGAAGCTCCCCCGGGCTGCTGG 1765
              |||||||       ||||||||||||||||||||   ||       ||||||||||||||||
Db   1618 -CTAGGACGACCGGGACCTGTCGCCGGGTACGCAGG 1676

Qy   1766 ACCAGCTCAATGCGCACGGTGGCGCCGGGGAACCCAAAAACACCGGTCGTCGTCATCAGACGGG 1825
              |||||||       |||||||||   ||||||||||||||||||||||||||||   ||||||||
Db   1677 ACAAGCTGATCTCGGCTGTCGCGACGCCAACCCGAACGACATCGTGGTCCTCAAACACG 1736

Qy   1826 GCACCCCCGAGGAGAGATGCCCTGGCTCTGCGACGCACGCCCCGTCATCCAGGCGCCTGGGTACG 1885
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db      -
```

```
Db   1737 GTTCGTCGGTCGGTGCTGATGCGGTGGCGTGTGTTCAAGACACCGGCGCCGCGGGCGTCCTTGGACATGTGGTACC 1796
              ||||| |||   || |||||||    |||       | ||     |||   ||||||||   |||||||||||||||
Qy   1886 GCCGCACGAGCACGGGCAAACTCCATTGCCGCAGCGTCGTTCGGCGGCCTGCGTGTTTCGGCACGCACAAACCCTTCGG 1945
              |||   |||||||||||||||||||||||||||||   |||||||||||||||||||||   |||||||   ||||||||
Db   1797 CGGGCCAGGCACGGGCGGGGGCCGAGGGCCCACCCGCGGGCCGGCTTCTACGGTGACGTCAACCCGAGCG 1856

Qy   1946 GCAAGCTGTCCCCTCAGCTTCCC----------------AAGCGCCTGCAGGACAACCCGCGTTTC 1996
              |||||||        ||||||||||               ||||||||||||||||||||   |||
Db   1857 GCAAGCTCACCGACAGAGTTTTCCCGGGCGCCGGGCGGCCGCGGACGCCGGAGTGGCGACCCGA 1916

Qy   1997 TCAAACTTCCGCACCGGAGCGCCGGGGCGGGCGGGGGGGGGGGTTACGGCGGGGAGGACGGTCTACGTTCGGGTACA 2056
              |||||       ||||||||||||||||||||   |||   |||||||   |||   |||||||||||||||||||||
Db   1917 CCAGCTACCCGGCGGTCGACAACCACGACGACGCTGCGGCCGCGCACCGTCGGGTACC 1976

Qy   2057 GGTACTACGGAGTTGCCGACCAGGACGGTCAATTGCCCTTTGGGCATGGCCGGGCTGCTCCTACA 2116
              |||   ||||       ||||||||||||||||||||   ||||||       |||   |||||||
Db   1977 GCTGGTTCGACAAAGGAGAGAACGTCAAGCCGCGGGTTCCCGGTTCCGGTTCCTGGGGCACGCCGA 2036

Qy   2117 CCACTTTTGCCTTTTCCAAATCTCTCCGTGTGTCTCACAAGGAC---GGCAAGCTGACGGTGT 2173
              |||       |||   ||||||||||||||   |||||||||||     ||||||||||||||
Db   2037 CCTCGTTCACGCAGAGCGCCCCCGACCGTTGCGGTAGCTCCACGGGTCGGTGGTCTCGAAGGTCA 2096

Qy   2174 CCCTTCTCCGTGAAGAAACAACCGCTCCGTTGCCCGGGCTGCGGCCCAGCTTCTTACGTCA 2233
              ||||||   ||||||||||   |||||||||   ||||||||||||||||||||   |||||
Db   2097 CGGTGCACGGTCCGCAACAGCCGGGGGAAGGCCGGCTGGCTCCAGGGGGGTACCTCG 2156

Qy   2234 AGCCCTCCAAGCGGCCCAAGAATTAACCGCCCGTTCAAGGAGCTCAAGGGCTTCGGCAAAGG 2293
              |||||||||||   |||||||        |||||||   |||||||||||||||||||||||
Db   2157 GTGCCAGCCCGGAACGTGACGGCTCCGCAGGCGGAGAAGAAGAGCTCGTGGGCTACACGAAGG 2216

Qy   2294 TCGAAACTGCAGCCCGGCCGAGACGAAGAGGGGGTGAC 2327
              || |||   ||||||||||||||||||||||||||||||
Db   2217 TCTCGCTCCGCGGGGCGGGGGCGGAGGGCGAAAGAGGGTGAC 2250
```

PS   Disclosure; Fig.14; 41pp; English.
XX
CC   The 0.75-kb PstI and the 3.05-kb Pst-NdeI fragments of C. biazotea
CC   ATCC 486 chromosomal DNA were sequenced using the dideoxy method. 8
CC   Sections of the coding strand were identified, the first 5 of which
CC   were contiguous (given in AAJ04785) and included a putative start
CC   codon. The sequences of the other 3, non-contiguous, downstream
CC   fragments are given in AAQ73991-93. Cloning of the gene region allowed
CC   extracellular prodn. of recombinant cellobiase in E. coli and S.
CC   cerevisiae.
XX
SQ   Sequence 1145 BP; 152 A; 444 C; 391 G; 158 T; 0 other;

Query Match   7.1%;   Score 177.4; DB 16; Length 1145;
Best Local Similarity 57.2%;   Pred. No. 1e-23;
Matches 362; Conservative   0; Mismatches 266; Indels   5; Gaps   2;

Qy   109 GGAGTCCCCTCTTCTCGCTTTACAGATGGCCCAACGGGGTAAGAGGGGACCAAGTTCTTC 168
          |||| || || ||||||| |||| |||| |||||| ||||||| ||||||||||
Db   448 GGCCTCGGCGAGGTCCGCCTGGGAGCTCCGGCGCGGGGCGCCGCCGGGGACCAAGTTCTCC 507

Qy   169 AATGGCGTCCTGCCTGCGGCCTGCCTTCCCTTGCGGCACGTGCCTTCCACATTCAAACCAA 228
          |||| || ||||||| |||| ||| ||||| |||| |||||||| ||| ||||| ||||
Db   508 GGCCGGCGCACCGTGCTGCGCTGTTCCCAACGCGACCCTGCCTCCGCGTGCCGTGGAGCGAG 567

Qy   229 ACTCTGCTCGAAGAGGCAGGTAAGATGAGATGGGCAAAGAGGCCATCGCTAAGAGTGCGGCAT 288
          |||| || || || |||||||||||||||||||||||
Db   508 GGCCGGCGCACCGTGCTGCGCTGTTCCCAACGCGACCCTGCCTCCGCGTGCCGTGGAGCGAG 567

Qy   229 ACTCTGCTCGAAGAGGCAGGTAAGATGAGATGGGCAAAGAGGCCATCGCTAAGAGTGCGGCAT 288
             |||||| || | | ||||||||| |||||| |||| ||||
Db   568 GAGTCGACGACCGAGGTCGGCGCCGCCGGCGTCTCGCCGAGGAGCCTCGCCAGCAGATCCAC 627

Qy   289 GTGATCCTCCGGCCCCGACTATCAACAATGCAAACGCTCCCCTCTTCGGTGGACGTGGCTTCGAG 348
          || | ||| | || ||||| ||| | |||| |||| |
Db   628 GTCCGTGCTCGGCCCCACGATCAACCTGACGCTCCACGGCTTCGGGGCGGCGACGCCCTGTTCGAG 687

Qy   349 TCGATTGGGTGAGGATCCCGGTTCCTCGGGGAGCTGCGGGGCTTGGGAGTTCCGGGTCTCATCCCGCCGCATT 408
          |||||||| || ||||| ||||| ||||| ||||||| || ||||||| ||||| ||||
Db   688 GCCTACTCCGAGGACCCGGCTGCTCACCGGCGGCCGGCGGGCGTGGGCGTGCGGCGCTGGGCGGCGGGCCTG 747

Qy   409 CAGAGACACTGGAGT-GCAGGCTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGGACAG 467
          ||| ||| ||| ||||||||||||||| || || || ||||||| |||
Db   748 CAGGAACCTCGGGCGTCGGCGGCGTCCCGCTCAAGGACCACCTGTCGCCAAACGACTCGGAGACCCA 807

Qy   468 GCGCATGATAGGTGCGAGAGCAATCGTCACGGAGGAGCGGGGCCTCTCCGTGAAAATCTACGGCCACTCCC 527
          ||||| | || |||| ||| | ||||| ||| |||||| ||||| |||||
Db   808 GCGCAACACCATGAAACTCCGTCGTCGACCCCCGCCGAGCTCTACCTCGTCC 867

Qy   528 GTTCCAGATTGCTGTTGCGAGACTCCCAGCCGCGGGTGCGTTCCATGACGGCGTACAATGGCAT 587
          |||||| ||| |||| ||| ||| ||| ||| |||| || || |||||||||| |||||||
Db   868 GTTTCGAGATCCCGGTCGACCAGCTCCGGTCCGTGGTCCGGTCATGGCCGCGGCTACAACGGACGT 927

Qy   588 CAATGGCGTGTCGTGCGGAGGAGAACCCTAAAATATCTTGATGGGAAGGGATGCTTCGGAAAAGGAATG 647
          | ||||| ||||| ||||| | | | | || |||| || |||| || |||||| ||| ||
Db   928 CAACGCGGTCCCCGCGAGCACCACCACGTGTCAACGAGCGTCCTCAAGGGCGAGTG 987

Qy   648 GGGTTGGGATGGCCTAATCATGAGGCGACTTGGTACGGCACATACTACGGAAGACGCCGT 707
          |||| | | ||| || | || ||||| || | || |||||| ||||| |||
Db   988 GGGGCTACACCGGCCTCGTCATTCGGACTGGTTCGGCACCCGGACCGCGCACCCGCC 1047

Qy   708 TGTGCGCAGGCGCTCGACCTCGAGATGCCCGGACC 740
          |||||||||| ||||||||||||||||||||||||
Db   1048 CG---CCGCCTCGACCTCGTGCTCATGCCCGGGGCC 1076

XX
CC   S. venezuelae macrolide beta-glycosidase gene desR, SEQ ID NO:23.
XX
DE   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..2340
FT   /*tag= a
FT   /product= "DesR"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999; 99WO-US14398.
XX
PR   26-JUN-1998; 98US-0105537.
XX
PA   (MINU) UNIV MINNESOTA.
XX
PI   Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77189.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 3; Page 369-370; 438pp; English.

CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryc gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics.
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAZ87286-Z87294 represent desosamine biosynthetic genes from
CC   Streptomyces venezuelae ATCC 15439, which encode proteins
CC   AAY77181-Y77189.
XX
SQ   Sequence 2430 BP; 407 A; 878 C; 834 G; 311 T; 0 other;

RESULT 6
AAZ87294
ID   AAZ87294 standard; DNA; 2430 BP.
XX
AC   AAZ87294;
XX
DT   05-JUN-2000 (first entry)
XX
DE   S. venezuelae macrolide beta-glycosidase gene desR, SEQ ID NO:23.

Query Match   7.0%;   Score 175.2; DB 21; Length 2430;
Best Local Similarity 47.6%;   Pred. No. 2.8e-23;
Matches 1016; Conservative   0; Mismatches 968; Indels 150; Gaps 11;

Qy   206 CGCTCGGTTCCACATTCAACCAAAACTCTCGCTCGAAGAGGCAGGTAAGAGATGATGGGCAAAG 265
          ||| ||||||| ||||||||| || ||| |||||||||| |||||||||| |||||
Db   353 CCCTGGCCAGCCACCTTCGACGACACCATGCCGGACAGCTTACGCGGGCAAGGTCATGGGCGCCGG 412

Qy   266 AGGCCATCGCTCGCTAAGAGTGCGCATGTGATCCTCCGGGCCGACTATCAACATGCAACGCGTCCC 325
             || | | || || |||||| | ||| | ||

```
Db    413 ACGGTCGCGGCGGCTCAACCAGGACAATGGTCCTGGGCCCGATGATGAACAACATCCGGGGTGC  472
Qy    326 CTCTGCGGTGGACGTTGGGCTTCGAGTCGGATCGGTTCCTTGGCGGCTTGGGAG          385
              ||  |||       ||   |  ||  |  || |||||||||||||||||
Db    473 CGCACGGCGCGGCGAACTACGAGACTTCAGCGAGGACCCCTGGTCTTCCGCGACCG          532
Qy    386 CTGCGGGCTCTCATCCGCGGGCCATTCAGAGCACTTGGACTACGAGCACTTTT          445
              |  |    |||     |||  |     ||  |   || ||||||||||||
Db    473 CGGTCGCGCCAGATCAAAGGGCATCCAGGGTGCGGGTCTTGATGACCACGCACAAGCTTCG    592
Qy    446 TGTGCAATGATCAGGAGGACAGGGCCGCCATGATGGTGCACAAGACATCTCACGGAGGGGCTC          505
              || | ||   |||||    ||||||||||||| |||||     |    ||
Db    593 CGGGCCAACAACCAGGAGAGAACAACCGCTTCTCCGTGAACGCCCAATGTCGACGAGCAGAGC    652
Qy    506 TCCGGTGAAATCTACGCACTCCCGTTCCAGATTGCCTGCGGACTCCCAGCCCGGGGGTCGGT          565
             ||||||   |||    ||||||    | | ||||||||||||||||||| |
Db    653 TCCCGGAGATCGAGTTTCCCGGCGTTCGAG--GGTGCTTCCAAGGCCGCGCCGCCCT          709
Qy    566 TCATGACGGCGTACAAATGGCCATCAAATGGCCGTTGTCGGCGAGAAACCCTAAATATCTTG          625
             |  |||| | ||||    ||  |||| || |||||||||||||||||||||||||
Db    710 TCATGTGTGCCTACAAACGGCCCTCAAACGGGGAAGCCGTCTCTGCGGCAACGAGCTCCTCA    769
Qy    626 ATGGGATGCTTCGAAAGGAAATAGGGGGTTGGGATGGCCTAATCATGAGCGACACTGGTACGGCA          685
             ||   ||||||||||||||||||||||||||||||||   |||    |    ||
Db    770 ACAACGTGGTGCTGCCCACGCAGTGGGGCGCTTCCAGGGCGTGATGCTCCGACTGGCTCGCCA    829
Qy    686 CATACAGTACCACAGAGAAGCCGTTGTTGTGGCAGGGCGGCTTCGAGACTTGCCCGGACCTCCAC          745
             ||  |||   ||     |    | ||||||||||||||||   |     ||||||||
Db    830 C---CCCGGCACCGACCGCCATCACCAAGGGCCTCGACCCTCCGGAGACGATG          873
Qy    746 GCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGGAGAAAAGCCCTTTATCCACGTCA          805
             |||||||||||||       |||  |||   ||||||||||||||||||| ||
Db    874 --------GGCGTCGAGCTCCCGGCGACGTCCCGACGGGCGAGCCGTCCGCCGCGGCCA          925
Qy    806 TTGACCAAGAGGCGGCTAGGGAAGTTCTTCAGTTCGTCAAGAGTGTGCTGCTCCCCTCGCGGAGTGA          865
             ||   |  ||||||||||||||||||||||||  |||||||||||||||||||
Db    926 AGTTCTTCGGCGA---GGGCGTGAAAGACGGCCCCGTTCCTGAAACGGCCACGGTCCCCGAGGGG    982
Qy    866 CGGAGAAGGCCCCCGAGACGACGACTGTCAACACACCCCCGAAAACGGCACGCTCTCCTCCGGA          925
             ||||||   ||||||||  |  |  ||||||||||||||||||||||||||||||||
Db    983 CCGTGACGGCGGGTCGGCGGGGACGGACGTCTGGGCCCAGATGGAAGTTCGGTCTCCTGCTCCTCG   1042
Qy    926 AGGTTGGGCAACGAGGAGGGCATCGTGCTGCTTGAAGAACGAGAACAACGTTCTGCCCTTGAGCA          985
             |||||||||||||||||||||   |||     |||||||||||||||||||||||||||
Db   1043 CCACTCCGGCGCCGCGGCGCGACAAGCGCGGTGCCCAACGCGGGTGTCCCGCAAGG   1102
Qy    986 AGAGAAGAGACCCTGATTGTGTCGGCCCCCAACGCCAAGGCGGCCACATACCACGGGCGGGGAG   1045
             |||       |||    ||     ||||||||||||||||||||||||||||||
Db   1103 TCGGCCGAGAAGGCGCGGCGGGCTGCTTCCTGCCCGAACGAGGGGGCCGCGCTCGCCGGGTG   1162
Qy   1046 GCTCTGCCGCCACTCAGGGGCCTACTACGCAGTGCACTCCCTTTGACGGGCCTCACGAGAGC   1105
             ||||||||||||||||       ||||||||||||||||||||||||||||||||||||
Db   1163 ACGGCGGCAAAGAGCATCGGGGTCATCGGCCCGACGGCCGTCGAC--CCAAGGTCACCGGC   1221
Qy   1106 TCGAGACGCCGCGCGCCGCCATCGTGTACACCGTCGGCGGCCGTTCCTCCCAATTCTAGGCG   1165
             ||||||||   ||||||||||     |||||||||||||||||||||||||||||||
Db   1222 CTGGGCACCGCCCACGTGTCTCGGCCGGCCGCCACTCGACACCATCACAGGCC   1281
Qy   1166 AGCAGTGCCTCACCCCCGACGACGGGCGCCTCCGGCCAGGCGGTCTTCAACGAGCCCC   1225
             |||       ||||||||||||||||||||||||||||||||||||||||||||
Db   1282 CGCGCGGGGTTGCGGGTGACGTAGCAGGGGTAGCGGGACCTTCGGACCGCAG   1341
Qy   1226 CTGGTACCCCTAACCGCGCACATTGACGACGCTCTTCTTCACCAAGAGGACACCACC   1285
             |||       ||||||||||||||||||||||    ||||||||||||||||
Db   1342 ATCCCGGCGGGGGGAACCTCAGC----------------CCGGCGTTCAACC   1375
Qy   1286 TGGTGGGACTACTACCACCCCAAGGCGGCAGACACGTGGATGGAGGGCACGT   1345
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1376 AGGGGCCACCAGCTCGAGCGCCAAGCGCGGTGTACGACGGCGCGTGACCGTGC   1435
Qy   1346 ACACCGCCGACGAGGACTCACCTACGACGCTCGGCGGCCTGTCCACGCGGTCCG   1405
             |||   ||||||||||    | ||||||||||||||||||||||||||||||
Db   1436 CCGCCGACGGCGAGTACCCGCACCCATCGGCGGGTGGTTACGGCCACAGGTGCAG-   1494
```

```
Qy   1406 CGTACGTAGACGACCAGCTTCGTCGTCGACAACGCCACCAAAGCAACGAGGTCCCCGGCGGATGCCT   1465
             |||||       |  |||    |||||||||||   |  ||   |||||||||||||||
Db   1495 --------------------------------------------CTCGGCCAGCCACA   1507
Qy   1466 TCTTTCGGCTTCCCCGCACCCGCGAGGAGACGAGGCGGCGCCATCAAATCTGCTTCAAGGGCAACAACGT   1525
             |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1508 CCATCGCGAGGCCGGTCAGGTTCTACGGGCACAAGGTGAGACGACCCGCGTCCTCAAGTCTGACCAAGG   1567
Qy   1526 ACAAGTTCAAGATCGAGTTGGCGCTTCCGGCACCTACGACCCTCACCACCTACAAGGGCGACACCATCG   1585
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1568 GCAACGCACAAGCT------------CACGATCTCGGGCTTCGCGGATGA   1603
Qy   1586 TCCCCGGCCCACGGCGGCTCCCTCCCGCCGGTCGCCAAGGTCATTGACGACCAGCCCGAAA   1645
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1604 GTGCCCACCCGCTCTCCCCTGGAGCTGGGTGACGGCGGGCGCGACCGA-   1663
Qy   1646 TCGAAAAGTTCCGTGCGCCCTCGCCCTGCCAAAGGAGCACGAGACCAGGTCATCACAGTCGCCGGGGCCTTA   1705
             ||   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1664 TCCGGAAGGCCGGTTGGAGTCGGCGCCGGAAGGGCGTACGGCGGTCGTTGTTCGC-------   1715
Qy   1706 ACGCCGACTGGGGAGACCGGAGGGCGGCGACCGGGCGGCGATGAAGCTCCCCGGGCGTGCTGG   1765
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1716 -CTACGACGACGGCACCGAGGGGCCGGTCGACCTTGTCCGGAACCTGTCGCTGCGGGGTACGCACGG   1774
Qy   1766 ACCAGCTCATTGCGGACCTGGCGGCCGGCGGCCGCGGCGAACCAAAAACCACGGGCGGCGACGGG   1825
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1775 ACAAGCTGAATCTCGGCTGTCGCGGTGCTCACGACGCCACGCACCCGGCGAACCACGATCGTGGCTGGACCG   1834
Qy   1826 GCACCCCGGAGAGATGCCCTGGCTCACAACGCCTGGCGGTCCGGCGCCGCGCCTCATCCTGGTACG   1885
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1835 GTTCGTCGGTTGCTATTGCCGTGTCAAGATGCCGTGGCTGTCCAAGACGCATGTGGGTACC   1894
Qy   1886 GCCGCAACGAGGACGGGCAACTCCATTGCCGACGTCGGCGCCCGCGCTGGTAACCCCTCGG   1945
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1895 CGGGCCCAGGCGGCGCCGAGGGCCGCCGAGGTCTCTTTGGCGGACTACTACACCCTCGG   1954
Qy   1946 GCAAGCTGTCCCTCAGCTTCCC---------AAGCGGCTGCAGGACAACCCCGCGTTTC   1996
             |||||||||||||||||||||         ||||||||||||||||||||||||||
Db   1955 GCAAGCTCACGCAGAGCTTCCCGGCGGCGGCTCTACGGCGGCTGGGCCCGGACCCGA   2014
Qy   1997 TCAACTTCCGCACCACGCGGGCGCGGGGCGCACGCGTTCTACGGCCGGGGGTACA   2056
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2015 CAAGCTACCCGGCGGTCGACAACCAGCAGGTGCCGGGCATCCACCGTGGGGTACC   2074
Qy   2057 GGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCTTTGGCCACGGCCTGTCCTACA   2116
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2075 GCTGGTTCGACAAGAGGAAGAACCGTCAAAGCCGCGCTGTTCCGGTTCACGCCTGTCTTACA   2134
Qy   2117 CCACTTTTGCCTTTTCCAAATCTCTCCGTGTCTGTCTCACAAGGAC---GGCAAGCTGAGCGGTGT   2173
             ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db   2135 CCTCGTTCACGCAGAGACAAGGCCCCCGACCGTCGTCGGTGGTC.TGAAGGTCA   2194
Qy   2174 CCCTCTTCCGTTGAAGAAGACAAGGGCCCGCTCCCTGCCCCAGGTGGGCCACTCTAGGTCA   2233
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2195 CGGTCACGGGTCCCCAACAGCGGGAAGCCCCAGGAGGTCGTTCGGTCCAGCCGTACCCTCG   2254
Qy   2234 AGCCCCTTCCACGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGG   2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||;|||
Db   2255 GTGCCAGCCCGAAACGTGACGGCTCCCGCCAGGGCAGGAGAGAAGCTCGGTGGGCTACACGAAGG   2314
Qy   2294 TCGAACTGCAGCCCGGCGACGGCGAAAGGCGGGTGAC   2327
             |||||||||||||||||||||||||||||||||||
Db   2315 TCTCGCTCGGCCCGGGGGGGGGGCGGGAAGACGGTGAC   2348
```

```
XX
DE   Streptomyces venezuelae DesR gene.
XX
KW   Glycosylated polyketide; modified recombinant bacterial host cell;
KW   mRBHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;
KW   tetracycline; polyene; polyether; ansamycin; isochromanequinone; sugar;
KW   desosamine; DesR; glucosidase; enzyme; gene; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             1..2430
FT                   /*tag= a
FT                   /product= "Streptomyces venezuelae DesR protein"
FT                   /transl_except= (pos:1..3, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO20229035-A2.
XX
PD
XX
PF   11-APR-2002.
XX
PR   05-OCT-2001; 2001WO-US31255.
XX
     05-OCT-2000; 2000US-238185P.
XX
PA   (MINU ) UNIV MINNESOTA.
PA   (LIUH/) LIU H.
PA   (SHER/) SHERMAN D H.
PA   (ZHAO/) ZHAO L.
XX
PI   Liu H, Sherman DH, Zhao L;
XX
XX
PT   Modified recombinant bacterial host cells in which the expression and
PT   activity of nucleic acids encoding sugar biosynthetic enzymes has been
PT   altered, useful for producing metabolites with altered sugar structures
-
XX
PS   Disclosure; Page 170-171; 174pp; English.
XX
XX
DR   WPI; 2002-405171/43.
DR   P-PSDB; AAE24237.
XX
CC   The invention provides a method to alter the sugar structure diversity
CC   for a particular metabolite via the recruitment and collaborative action
CC   of sugar genes from a variety of sugar biosynthetic pathways to yield a
CC   metabolite comprising a non-natural sugar, e.g., a novel glycosylated
CC   polyketide. The invention also relates to a modified recombinant
CC   bacterial host cell (mRBHC) in which the expression and activity of
CC   nucleic acids encoding sugar biosynthetic enzymes has been altered.
CC   The mRBHCs may be cultured to produce the modified sugar products,
CC   e.g. a macrolide, anthracycline, angucycline, avermectin, milbemycin,
CC   tetracycline, polyene, polyether, ansamycin or isochromanequinone.
CC   The present sequence is Streptomyces venezuelae sugar (desosamine)
CC   biosynthetic gene cluster DesR (glucosidase) gene.
XX
SQ   Sequence 2430 BP; 407 A; 878 C; 834 G; 311 T; 0 other;

Query Match             7.0%;   Score 175.2;  DB 24;  Length 2430;
Best Local Similarity   47.6%;  Pred. No. 2.8e-23;
                        Matches 1016;  Conservative  0;  Mismatches  968;  Indels 150;  Gaps 11;
```

```
Qy    206 CGCTCGGGTTCCACATTCAACCAAACTCTCTCGAAAGAGGCAGGGTAAGAATGAATGGGCAAAG 265
Db    353 CCCTGGCCAGCACCTTCGACGACACCATGGCCGACAGCTACGGCAAGGTCATGGGCCCGCG 412

Qy    266 AGGCCATCGCTTAAGAGTGCGGCCATGGTGATCCCTCGGCCGCTATCAACATGCAAGCTCCC 325
Db    413 ACGGTGCGCGCCGCCTCAACCAGGACATGGCTCCTGGGCCCGATGAATGAACAACATCCGGGTGC 472

Qy    326 CTCTCGGTGGACGTGGCCTTCGAGTTGGATGGTTGACGATCCGTTCTCGGCGGCTTGGGGAG 385
Db    473 CGCACGGCGCGGCCGGAACTACGAGACCTTCAGCGAGGAGAGACCCCCTGGTGTCTCGGCGCACGG 532

Qy    386 CTGCGCGCTCTCATCCGCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGGATCAAGCACTTTT 445
Db    533 CGGTCGCCCCAGATCAAGGGCCATCAAGGGGTCGGGCTCGATGGACCACGGGCCAAGCACTTCG 592

Qy    446 TGTGCAAATGATTCAGGAGGGACAGGGGCATGATGGTGCAGAGCATCGTTCACGGAGGGGCTC 505
Db    593 CGGCCAACAACCAGGAGAACAACCCGCTTCTCCGGTGAACCGCCAAGTCGACGAGCAGACGC 652

Qy    506 TCCGGTGAAATCTACCGCACTCCCGTTCCAGATTGCTCGTGTGGCGAGACTCCCACGCGGTCGT 565
Db    653 ACAACGTGCGTCGCACGCACTGGGCGCTTCCAGGGCGTGGGTGATGTCGACTGGCTCGCCA 709

Qy    566 TCATGACGGGGTACATCACAATGGCGGTTCCCGGCGTTGCAGC---GGCGTCCTCCAAGGCCTGCACGCG 625
Db    710 TCATGTGTGCCTACAACGGCCTCAACGGGGAAGCGGCCGGCGGGAAGAGCCGGCGGAGGTCCTCA 769

Qy    626 ATGGGGATGCTTCGAAAGGAATAGGGGGTTGGGGATGGCCTAATCATGCGACGACTGGTTACGGCA 685
Db    770 ACAACGTGTCGCGCACGCACTGGGCGCTTCCAGGGCTGGGTGATGTCGACTGGCTCGCCA 829

Qy    686 CATACAGTACCACACAGAAAGCCGTTGTGGCAGGCCTCGACGATGCCCGGGAACCTCCAC 745
Db    830 C---CCCGGGCCACCGACCGACCGCCATCACCCAAGGGCCTTCGACCAGGAGATG------- 873

Qy    746 GCTTCCGAGGAGAAAACACTCAAGTTCAACCTCTCCAAACGGAAAGCCCTTTATCACGTCA 805
Db    874 ------GGCGTCGAGCTCCCCGGCGACCTTCCCGAAAGGCGCCGCCCTCACCGGCA 925

Qy    806 TTGACCAGAGAGGGCTAGGGGAAGTTCTTCAGTTCGTTCACGGCAGGCTGGCGGAGTGA 865
Db    926 AGTTCTTCGGCGA---GGCGCTGAAGACGACCGCCGGTCCTGAAGCGGCACGGGTCCCGAAACGGCG 982

Qy    866 CGGAGAACGGCCCCGAGACGGACGACTGTCAAACAACACCCCGAAAACGGCCAGCTCTCCTCCGGA 925
Db    926 CCGGTGACGGGGTCGCGGGCGACGGGATCGTCGGCGGAGGATCGAGGATGCCCGGGACCTCCAC 985

Qy    926 AGTTCTTCGGCGA---GGCGCTGAAGACGGACCGCCGGTCCTGAAGCGGCACGGGTCCCGAGGCGG 982
Db    983 CCGTGACGGGGTCGCGGGCGACGGGATCGTCGGCGGAGGATCGAGGAAGTTCGGTGTCTTCCTCG 1042

Qy    986 AGAAGAAAGAGACGCCTGATTGTTGTCGGCCCCCAAACGGCCAAGAGCGGCCACATACCACGGCGGGAG 1045
Db   1043 CCACTCGGGGCGGGCCCCGGAGCGGCGACAAGAGGCGGGGTGCCCAGGCGGGTGTCCCCAAG 1102

Qy   1046 GCTCTGCCCGGACTCAGGGGCCTACTACGGCCAGTCACTCCCCTTTGACGGGCCTCAGCGAAGCAGC 1105
Db   1103 ACGCCCGGCGGCAAGAGCAATCGGGTGCATCGGGGCCCCTGCCCGGGTGGCCAGGCCTGCCCGGGTG 1162

Qy   1106 TCGAGACGCGCCCATCGTACACGTCGGCCGCGCCCTACACCACCGGTTCCTCCCATTCTAGGCG 1165
Db   1163 ACGCCCGGCGCAAGAGCAATCGGGTGCATCGGGGCCCCTTGACGGGGCTCAGCGCCCGGTG 1221

Qy   1166 AGCAGTGCCTCACGCCCACCCGACGGCGGGCGGCCCACTCGACACCATCAAGGGC 1225
Db   1222 CTGGGCCAGCGCGCCCACGTCGGCCCCACGGGCGGCCACGTTCCTCCCATTCTAGGCG 1281

Qy   1226 CTGGTACCCCTTAACCGCCACGACCATTGAGCACGGGCTGCTTCTTCACCAAGAGGGACATGCACC 1285
Db   1282 CGCCGGGGTCGGGGTTCGGGGTGACGTACGGAGACGGGGTCACGACCTTTGGGGCACGAGCCCC 1341

Qy   1286 TGGTGGACTACTACACCCCAAGGGGGCACACGTGGTACGGCCGACATGGGAGGGCACGT 1345
Db   1342 ATCCCCGGGCGGGGGAACCTCAGC-------------------CCGGCGTTCAACC 1375

Qy   1346 ACACCGCGCCGACGAGGACTGCACCTACGAGCTCGCGGCCTCGGCCGCGGGGACCACAGGGCAAAGG 1405
Db   1376 AGGGCCACCAGCTGACGGCGGGGCAAGGGCGGCGGGGCACGGCCTGCACCGTG 1435

Qy   1406 CGTACGTAGACGACCAGCTCGTCGTCGACAAACGCCACCAAGCAGGTCCCCGGCGCGGATGCCT 1465
Db   1436 CCGCCGACGCGCGAGTACCGGCTCCGTGCCACCGGTGGTTACGCCGCACGGTGCAG- 1494

Qy   1406 CGCACGGCGCGGCGGAACTACGAGACCTTCAGCGAGGAGACCCCCTGGTGTCTCGGCGCACGG 1465
Db   1495 -----------------------------------------------CTCGGCGAGCCACA 1507
```

```
Qy 1466 TCTTCGGCTTCCGCCACCCGGAGGAGGACGGGGCCGGCCATCAAATTCGTCAAGGGCAACACGT 1525
             |||             |||||     |            |||   |||       |||
Db 1508 CCATCGAGCCGGCGGTCAGGTCTACGCACAAGGTGAGCAGCGCGTCCTCAAGCTTGACCAAGG 1567

Qy 1526 ACAAGTTCAAGATCGAGTTCGGCTTCCGCCACCACCTACACCCTTCAAAGGGCGACACCATCG 1585
             ||||||       ||               |||||||||||||||
Db 1568 GCACGCACAAGCT----------------------CACGACTCTCGGGCTTCGCGATGA 1603

Qy 1586 TCCCCGGCGCACGGCGTCCTCCCTCCGGGCTGCGCCTGCAAGGTCATTGACGACGCGAAA 1645
                                              |||||       ||||||
Db 1604 GTGCCACCCCGGCGTCTCCTGGACTGGCGCCTGGCGTTCACGGCGCACGGCGACGA 1663

Qy 1646 TCGAAAAGTCCGGCTGCCCTTCGCGCAAGGAGCACGACCAGGTCATCACGCGGGCCCCTTA 1705
             |||||||||           |||        |||||           |||||||
Db 1664 TCGGGAAGGCCGTGGAGGTCGGGCTCGCCCGCGGGAGGCCGTAGCGCGGTCGCTTCGC----- 1715

Qy 1706 ACGCCGACTGGGGAGACCAGAGGGCCCGACCGGCGCGATGAAGCTCCCCGGCGGTGCTGG 1765
             ||||      |||||||           |||       |||||       ||| ||
Db 1716 -CTACGACGACGGCCACCCGAGGGCGTCCGACCGGTCCGCTCGGGGTTACGCGAGG 1774

Qy 1766 ACCAAGCTCATTGCCGACGCTGGCCCGCGGGGACCAAAACACGTTCGTCGTCATGCCAGACGG 1825
             ||||     |||||         |||           ||||||   || |||||
Db 1775 ACAAGCTGAATCTCGGCTCTGCTGCGGACGGCCAACCCGGACCGGTCCTCAAACACG 1834

Qy 1826 GCACCCCGGAGGAGATGGCCCTGGCTCGACGCCACGCGCCGCTGCATCCAGGCCCTGGGTACG 1885
             ||       ||||||||           |||||||||||   |||||||||||||||||
Db 1835 GTTCGTCGGGTGCGTGATGCCGGTGTTCCAAGACACCGGCTGGCTGGGGCGGGTGTCCGTGTACC 1894

Qy 1886 GCCGCAACGAGACGGGCAACTCATTGCCGACGTCGTCGTTTGGCGACTACAACCCCTCGG 1945
             |||||||||||||||||       ||           |||         ||||||
Db 1895 CGGGGCCAGGCGCGGGCCCACCACCGGCGTGCTCGTTACGGTGGACGTCAACCCGAGCG 1954

Qy 1946 GCAAGCTGTCCCTCAGCGTTCCC-------------AAGCGCCTGCGAGGACAACACCCGCGTTTC 1996
             ||||||||||||||       |||||           ||||||||
Db 1955 GCAAGCTCACGCAGACGTTCCCGGCGGAGAACCACCACGCGGTCGCCGCCGCCACCCGA 2014

Qy 1997 TCAACTTCCGACAACGGCTGTACGCGGGGGACGACTGGCTACGTGCTCGGGGTACA 2056
             ||||||||||||||||||||       ||||||||||||||||||||||||||
Db 2015 CAAGCTACCCGGGCGTCGACAACCAGACGTACCCGCGAGGGCATCCACGTCGGGTACC 2074

Qy 2057 GGTACTACCGAGTTTGCCGACAAGGGACGGTCAAATTCCCCTTTGGCCACGGCGCCTGTCCTACA 2116
             |||||||||           |||||||         |||||         |||||||||||||||||
Db 2075 GCTGGTTCGACACAAGGAGAACCGTCAAGCCGCGCTTCCCGTTCGGCTGCCCGTGCTGTCCTACA 2134

Qy 2117 CCACTTTTGCCTTTTTCCAATCTCTCCGTGTCTCACACAAGGAC----GGCAAGCTCGAGCGTGT 2173
             |||       ||           |||||||||||||||||       ||||||||||||||||
Db 2135 CCTCGTTCACGCAGGACGGCCCCGACCGTCGTGGCTACGTTCCACGGGTCGTCGAAGGTCA 2194

Qy 2174 CCCTCTCCGTGAAGAACACCAGCGGCTCGGTGCCCCAGCTCTACGGTCA 2233
             |||           ||||           |||||           |||
Db 2195 CGGTCACGGTCCCCAACACGGGGAAGCGCCCGCGCGGTCGTCCAGGGGTTACCTCG 2254

Qy 2234 AGCCCTCCAAGCGGCCAAGATTAAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAG 2293
             ||||||           |||           |||           ||| |
Db 2255 GTGCCAGGCCGAACGTGACGGCGTCCGGAGCGAAGGGGGTGAC 2314

Qy 2294 TCGAACTGCAGCCCGGCGAGACGAAGGGGGTGAC 2327
             ||||||           |||||||||||||||||
Db 2315 TCTCGCTCCGCCGCGGGGGCGGAAGACGGTGAC 2348
```

KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO20000006200-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999; 9WO-US14398.
XX
PR   26-JUN-1998; 98US-0105537.
XX
PA   (MINU) UNIV MINNESOTA.
XX
PI   Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77179.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 2; Page 281-287; 438bp; English.
XX

CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active
CC   monomers. The compounds are useful
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the desosamine
CC   biosynthetic gene cluster from Streptomyces venezuelae ATCC 15439.
XX
SQ   Sequence 12441 BP; 1704 A; 4294 C; 4686 G; 1757 T; 0 other;

RESULT 8
AA287284
ID   AA287284 standard; DNA; 12441 BP.
XX
AC   AA287284;
XX
DT   05-JUN-2000 (first entry)
XX
DE   S. venezuelae desosamine biosynthetic gene cluster pikB, SEQ ID NO:3.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;

```
Query Match             7.0%;  Score 175.2;  DB 21;  Length 12441;
Best Local Similarity   47.6%;
Matches 1016;  Conservative     0;  Mismatches 968;  Indels 150;  Gaps 11;

Qy   206 CGCTCGGTTCCACATTCAACCAAAACTGTCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 265
             ||||| ||||||||||    ||||   ||           |||         || ||||||
Db  3977 CCCTGGCCAGCACCTTCGACGACACCATGGCCGCCGACGACTATCACATCACGGCTCCC 4036

Qy   266 AGGCCATCGCTAAGAGTGCCGCATGTGATCCCTCGGCCCCGACTACGGCCAAGGTCATGGGCGCCCG 325
             |||     |||||||||||||         || ||||||||||||||||||||||||||||||||||
Db  4037 ACGGTCGCGCGTCAACCAGGACATGGTCTTCCTGGGCCCCGATGATGACAAACATCCGGGTGC 4096

Qy   326 CTCTCGGTGGACGTGGCCTTCGAGTCGATTGGTGTGAGGGATCCGTTCCTCGGCCGGGCTTGGGAG 385
             |||        |||||||||||||||||||||||     |||       ||||||||||||||||||
Db  4097 CGCACGCGCGCCGGAACTACGAGACCACCTACGGGAGGTGCAGCTACGGAGATCAAGCACTTTT 4156

Qy   386 CTCCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGGTGCAGCTACGGAATCAAGCACTTTT 445
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4157 ...
```

Db 4157 CGGGTCGGCCCAGATCAAGGGCATCCAGGGTGCGGGCTCTGATGACCACGGCCACTTCG 4216
Qy 446 TGTGCAATGATCAGGAGGACAGGGGCATGATGGTTGGCAGAGCAATGGTGTCACGGAGGCGGGGCTC 505
Db 4217 CGGGCCAACAACCAGGAGAACCGGCTTCTCCGTGAACGCCAATGTCGACGAGCAGCAGACGCCGAAA 4276
Qy 506 TCCGTGAAATCTACGCACTCCCGTTCCCAGATTGCTGTCGACTCCCAGCGCGGGTGCGGT 565
Db 4277 TCCCGCGAGATCGAGTTCCCGCGTTCGAG---GGGTCCTCCAAGGCCGGCGCGGCCTCCT 4333
Qy 566 TCATGACGGCGTACAATGGCATCAATGGGGTGTCGTGCAGCGGCGAGAACCCTAAATATCTTG 625
Db 4334 TCAGTGTGCCTCACAACGGCCTCAAACGGCGTCCTGCGCGAAACGACGACGACTCCTCA 4393
Qy 626 ATGGGATGCTTCGAAAGGAAATGGCGGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCA 685
Db 4394 ACAACGGTCGCACGGCACGCAGTGGGGCTTCCAGGCCTGGGGTGATGTCCAGACTGGCTCGGCCA 4453
Qy 686 CATACAGTACCACGAGAAGGCCGTTGTGGCAGGCCTCGACCTTCGAGATGCCCGGACCTCCAC 745
Db 4454 C---CCCGGGCACCGACGCCATCACCAAGGGCTCTGACCAGGAGATG--------- 4497
Qy 746 GCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAAGCCCTTTATCCACGTCA 805
Db 4498 -------GGCGTCGAGCGTCCCCGGCGACGTCCGGAAAGGGCGACCCCTCGCGCGCGGCCA 4549
Qy 806 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTTCGTCAAGAAGTGTTCGGCCTCCGGGAGTGA 865
Db 4550 AGTTCTTCGGCCGA---GGGCCTGAAGACGGCCGGTCCTGAACGGCACAGGTCCCCGAGGGCG 4606
Qy 866 CGGGAGAACGGCCCCCGAGAGACTGTTCAACAAACACCCCGCAGCTCTCCTCCGGA 925
Db 4607 CCGTGACGGCGCGGTCGGCGGGGGGATCGTGGGCGGAGAAGTTGGGTGCTCGGGTCTCG 4666
Qy 926 AGGTTGGCAACGGCATCGTGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCA 985
Db 4667 CCACTCCGGGGCGGGCCGCGACAAAGGGCGGGTGTCCCGCAAGG 4726
Qy 986 AGAAGAAGAAGACGCTGATTGTCGGGCGCCCCAACGCCACGCGGGGAG 1045
Db 4727 TCGCCCGAGACGACGGGGGGGTCCTGCGCCAACGAGGGCCCTGCCGTGGGGTG 4786
Qy 1046 GCTCTGCCGCACTCAGGGCCTACTACGCACTCACTCCCTTTGACGGGCCTCAGCAAGCGGC 1105
Db 4787 ACGCCCGCAAGAGATCGCGGTCATCGCCGACGGCGGTCGAC-CCCAAGGTCACCGGC 4845
Qy 1106 TCGAGACGGCCCCCCATCGTACACCGGTTCGGCGGCCTACACCACCACGTTCCTCCCATTCTAGGCG 1165
Db 4846 CTGGGCAGCGCGCCCAGTCGTCGTCCCCGGACTCCGCGCCACTGCACCATCAAGGCC 4905
Qy 1166 AGCAGTGCCTCACGCCCGACGCCGGGGGCATGCCGCTGGAGGGGTCTTCAACGAGCCCC 1225
Db 4906 CGCGCGGGGTGGACGGTGACGTACGAGACGGGTGAGGACAACCTTCGGGACGGAC 4965
Qy 1226 CTGGGTACCCCTTAACCCCCGACACATTGACCGAGCTCTTCTTCACCAAGACGGACACC 1285
Db 4966 ATCCCGCCGGTGGGGAACCTCAGC---------CCGGCGTTCAACC 4999
Qy 1286 TGGTTGGACTACTACCACCCCAAGGCGCGCAGACACGTGGTACGGCCGACGCACGT 1345
Db 5000 AGGGCCACCAGCTCGAGCGGGGGCAAGGCGGGGCGGGGGGCTGTAGGCGCACCGCTGACCGTGC 5059
Qy 1346 ACACCGCCGCGACGGGACTGCACCTACGGACTCTCGGCCTGCTCGCTGCTGCGGGACGGCAAAAG 1405
Db 5060 CCGCCGACGGCGGAGTACCGCATCGGCGGTCCGTGGTTACGGCCACGGTCCAG- 5118
Qy 1406 CGTACGGTAGAGGACCAGCTCGTCGTGTGGACAACGCCACCACCAGCAGGTCCCCGGCGATNGCCT 1465
Db 5119 ---------------------------------CTCGGCAGCCACA 5131
Qy 1466 TCTTCGGCCTCCGCCCACCCGCGAGGAGACGGGCGCGCATCAATCTCGTCAAGGGCCAACACGT 1525
Db 5132 CCATCGAGGCCGGTCAGGTGTCTACGGCAAGGTGAGCAGCACCGCGCTCCTTCAAGCTGACCAAGG 5191

Qy 1526 ACAAGTTCAAGATCGAGTTCGGGTCTCCGACCACCGCGACGAAGGGCGAGACACCATCG 1585
Db 5192 GCACGCACAAGCT---------------------CACGATCTCGGGCGCTTGCCCGATGA 5227
Qy 1586 TCCCCGGGCCACGGCGTCCCCTCCCGCGCTGGGCGGCGCGCGGGGGTCGGGGGCTGGCTGCAGCGGTCATTGACGACCAGGGGGGGA 1645
Db 5228 GTGGCCACCCGCGCTCCCCTGGGACGTCTGGGGCGTGGGTGACCGCGGCGGCGAGCGGCAGGACGA 5287
Qy 1646 TCGAAAGTCCGTCGCCCCTTCGCCAAGGACGACCAGGTTCATCATCTTGCGCGGGGGCCTTA 1705
Db 5288 TCGGCGAAGGCGCGTGGAGTGCGGCGGGGGAAGGGCCCGTACGGCGGCGTTCCGTGCTGG- 5339
Qy 1706 ACGGCCGACTGGGGAGGGACCCGCGGAGGGCGCGGGCGGGCGGCGGCATGGAAGCTCCCGGCCTGCTCGG 1765
Db 5340 -CTACGACGGACGGCGCACCGCAGGGGCTCGACCCTCCGGAACCTGTCGCTGGCTGCGGGGGGTACGCAGG 5398
Qy 1766 ACCAGCTCATTGCCGACGTGGGCCGGGGCGGGGAACCCAAAACACCGGTGGGTCGGTCGTCATGCGACGGG 1825
Db 5399 ACAAGCTGATCTCGGCTGTTGCGGCGACGGCCAACCCGAACACGATCGGTGGTCCTCAAACACCG 5458
Qy 1826 GCACCCCCGAGGAGAGATGCCCGGCTCGAGCGACGCACCAGGCCCGCGGTCATCCAGGCCTGGTACG 1885
Db 5459 GTTCGGTCGGGTGGCTGACAAGGGGCGCGGGCTCTCCTTGGCGGCGGGGTCCTTGGACTACAGCTGG 5518
Qy 1886 GCGGCAACGAGACGGGGCAACTCCATTGCCGACGGTCGTCTTTGGGCGGGGGTCGCGTGT 1945
Db 5519 CGGGGCCAGCGGGGGGCGGGGGCCACCGGCCGACCGCCTGCTCTACGGGTGACGTCAACCGGAGCG 5578
Qy 1946 GCAAGCTGTCCCTCAGCTTCCCC-----------AAGCGCCTGCAGGACAACACCCCCGTTTC 1996
Db 5579 GCAAGCTCACGGCAGAGCTTCCCGCCGGCCCCGAGGGGCGGGCCGGGAGGCGGGGAGGAGACGGACCCCGA 5638
Qy 1997 TCAACTTCCGCACCGAAGGCGCGGGGGGCACGCGGGACGTGTAGGGCGGGGACGTCTACGTCGGGGTACA 2056
Db 5639 CAAGCTACCCGGCGTCGACAAGCAGCAGACGTACGGCGGGCATCCACGTCGGGGTACC 5698
Qy 2057 GGTACTACGAGTTTGCGACAAGGACGTCAAATTTCCCTTTGGCCACGGTCGTGCTACA 2116
Db 5699 GCTGGTTCGACAAGGGAGAACGTCAAGGCCGTTCCCGGTTCCCGGTTCTGCGTCGTACA 5758
Qy 2117 CCACTTTTGCCTTTTCCAATCTCTCCGGTGTCTCACAAGGAC---GGCAGCTGAGCGTGT 2173
Db 5759 CCTCGTTCACGCAGGAGCCCCGACCGTCGTGCGTACGTCCACGGTCA 5818
Qy 2174 CCCTCTCCGTGAAGAAACACCGGGCTCCGTGCCGGGCGCACAGGTGGCCGACAGCTTCACGGTCA 2233
Db 5819 CGGTCAACGGTCCGCAACACGGGAAGCGGGGGCCGGCGCCGGAGGTGTTCCAGGGGGTACCTCG 5878
Qy 2234 AGCCCCTCCAAGCGGACGATTAACCCGCCGGTCAAGGGACTCAAGGGCGGCAAAGG 2293
Db 5879 GTGCCAGCCCGAAACGTGACCGGCCTCCGGCCAGGCGAAGAGAGAGCTCGTCGGCTACACGGAAG 5938
Qy 2294 TCGAACTGCAGCCCCGGCGACGAGAGGGCGGTGAC 2327
Db 5939 TCTCGCTGCCGCCGGCGGCGGCGAGGCGAGCGAAGACGGGTGAC 5972

```
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..809
FT                   /*tag= a
FT                   /product= "Streptomyces venezuelae Des gene cluster
FT                   encoded protein #1"
XX
PS
CDS                  806..2014
FT                   /*tag= b
FT                   /product= "Streptomyces venezuelae Des VIII protein"
FT                   /transl_except= (pos:806, 808, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             2162..3442
FT                   /*tag= c
FT                   /product= "Streptomyces venezuelae Des VII protein"
FT                   /transl_except= (pos:3532,3534, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (3535..4245
FT                   /*tag= d
FT                   /product= "Streptomyces venezuelae Des VI protein"
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement 4312..6741
FT                   /*tag= e
FT                   /product= "Streptomyces venezuelae Des R protein"
FT                   /transl_except= (pos:4312..4314, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (7969..6829)
FT                   /*tag= f
FT                   /product= "Streptomyces venezuelae Des V protein"
FT                   /transl_except= (pos:7969..7967, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (8979..7966)
FT                   /*tag= g
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT                   /transl_except= (pos:8979..8977, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (9111..9989)
FT                   /*tag= h
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (11639..10182)
FT                   /*tag= i
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT                   /complement (12883..11636)
FT                   /*tag= j
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT                   /transl_except= (pos:12881..12883, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
```

PT   activity of nucleic acids encoding sugar biosynthetic enzymes has been
PT   altered, useful for producing metabolites with altered sugar structures
PT   .

Disclosure; Fig 8; 174pp; English.

The invention provides a method to alter the sugar structure diversity
for a particular metabolite via the recruitment and collaborative action
of sugar genes from a variety of sugar biosynthetic pathways to yield a
metabolite comprising a non-natural sugar, e.g., a novel glycosylated
polyketide. The invention also relates to a modified recombinant
bacterial host cell (mRBHC) in which the expression and activity of
nucleic acids encoding sugar biosynthetic enzymes has been altered.
The mRBHCs may be cultured to produce the modified sugar products,
e.g. a macrolide, anthracycline, angucycline, avermectin, milbemycin,
tetracycline, polyene, polyether, ansamycin or isochromanequinone.
The present sequence is Streptomyces venezuelae sugar (desosamine)
biosynthetic gene cluster.

```
SQ   Sequence 13613 BP; 1858 A; 4732 C; 5093 G; 1930 T; 0 other;

Query Match             7.0%;   Score 175.2;   DB 24;   Length 13613;
Best Local Similarity  47.6%;   Pred. No. 3.1e-23;   Indels 150;   Gaps 11;
Matches 1016;   Conservative   0;  Mismatches   968;

Qy    206 CGCTCGGTTCCACATTGCAACCAAAACTCTGCGTCGAAGAGGCAGGTAAGATGATNGGCAAAG 265
          |||||||||| ||   |||  ||| |||||||  ||||||| |||||||||||||| |||||
Db   4664 CCCTGGCCCAGAGACCTTGCGCACGCAGTGGGGCTTCGACCTCGGGGTTGGGCCCGG 4723

Qy    266 AGGCCATCGCTAAGAGTGCGCARAGTGATCCTTCGGCCCCGACTATTCARACATGCAAACGCTCCC 325
          |||||||||||||||||||   || || |||  ||||||||||||||||||||||||  |||
Db   4724 ACGGTCCCGCGCGTCAACCAGACATGGTCTGGCCCCGATGATGAACAACATCGGGGTGC 4783

Qy    326 CTCTCGGTTGGACGTGGCGTTCGAGTCGCATTGGTTGGAGGGATNCCGTTCCTGGCCGGGCTTGGGAG 385
          | ||||||||||||||||||||| || |||||| || |||| ||| ||| ||||||| ||||||
Db   4784 CGCACGCGGCGCGGGAACTACGGAAGCCTTCAGCGAGGAGACCCCCTGGTCTCCTCGCGGCACCG 4843

Qy    386 CTGCGCGGCTCTCATCCCGCGGCATTCAGAGCACTGGACTGCAGGCTACGATCAAAGCACTTTT 445
          |||||||| ||||||   ||   || || || || || || ||||||||||||  |||||||||
Db   4844 CGGTCGCCCAGATCAAAGGGCATCCAGGGTGCGGGGTCAGTAAGCGCCAAGCACTTCG 4903

Qy    446 TGTGCAATGATCAGGACGACAGGCGCCATGATCGACAGCCATCTTCACGGAGGGGGCTC 505
          |||| || |||| || |||| ||||| ||  ||||||||| |||||||| ||||||||
Db   4904 CGGCCAACAACCAGGAGAAACAACCGGCTTCTCCGTGAACGGCAAATGTCGACCAGCAGACGC 4963

Qy    506 TCCGTGAAAATCTACGGCACTCCCGTTCCAGATTGCTCGTGCGAGACTCCCAGCGGGGTGCGT 565
          ||| || || |||| || ||  ||||||||||||| ||||||| ||||| ||||||||||||
Db   4964 TCCGCCAGATCGAGTTYCCCGCGCGTTCGAG---GCGTCCTCCAAGCGCCGGGCGCGCCTCCT 5020

Qy    566 TCATGACGGCGTACAAATGGCATCAAATNGGCGGTGTCGTGCAGGGAGAACCCTAAATATCTTG 625
          ||||||||| ||||  ||| ||| |||  |||||||||||| ||||||||   |||||||||||
Db   5021 TCATGTGTCGCCTACAACGGCCTCAACGGAGGAAGCCGTTCCCACGGCACAGACGAGCTCCTCA 5080

Qy    626 ATGGGATNGCTTCGAAAGGAAATGGGGTTGGGATGGCGCTAATCATGAGCGACTI3GTACGGCA 685
          ||||||| ||||||||||||||||||| || |||| ||||||||||||||||||||||||||||
Db   5081 ACAACGTGCTCGCGCACGCAGTGGGGCTTCCAGGGTGGGTGGGTGCGACTGGCGTCGCCA 5140

Qy    686 CATACAGTACCACAGAAGCCGTTGTGGCAGGCCTGACCTCGAGATNGCCC3GACCTCCAC 745
          |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db   5141 C---CCCGGGCACCGACCGCCATCACCCAAAGGGCCTCGACCAGGAGATG------------ 5184

Qy    746 GCTTCCGAGGAGGAAACACTCAAGTTCAACGTGTCTCCAACGGAAAGCCCTTATCCACGTCA 805
          |||    ||||||||| ||||| ||| || ||| ||   ||  |||||||| |||||||
Db   5185 -------GGCGTCGAGCTCCCCGCGACGTTCCCGAAAGGGCGAGCCCTGTCGCCGGCGGCCA 5236

Qy    806 TTGACCAGAGGGGCTAGGGAAGGTTCTTCAGTTCGTCAAGAAGAAGTGTTGCTCGCTCCGGAGTGA 865
          ||||||||||||||||||||||||||||||||||||||   ||||||||| |||| || ||||
Db   5237 AGTTCTTCGGCCGA---GGCGCTGAAGACGACGCCGTGCTCGAAGCGGCACGGTGTCCCAGCGG 5293

Qy    866 CGGAGAACGGCCCCGAGACGGACTTGCAAACAACACCCCGAAAACGGCAGGTCTTCCTTCCGGA 925
          |||   ||| || || |||| ||| ||   || |||| |||  || |||  |||| ||||||||
Db   5294 CCGTGACGCGGTCGGCGGAGCGGAGATCGGTCGGCCCAGATGGAGAAGTTCGGTGTCTGCCTCG 5353
```

```
Qy    926 AGGTTGGCAACGAGGAGCATCGTGCTGCTGAAGAACGAGGAGAACAACGTTGTGCCCTTGACCA 985
              || |  ||  |||||     |||   |||  || ||    | | | |||   |||||||||
Db   5354 CCACTCCGGCGCGCCGGGCCCGACGCGCCGGGTTGCCCAGGCGGGTGTCCCGCAAGG 5413

Qy    986 AGAAGAAGAAGAGCCTGATTGTGCGGCGCGACAAGCGCGGGTTGCCCAGGCGGGAG 1045
Db   5414 TCGGCGAGGAGAACGGCGCGGTTGCGTCCTGCGCAACGACGCCACGGCGCGGTG 5473

Qy   1046 GCTCTGCCCACTCAGGCGGCTTACTACGGAGTCACTCCCTTTGACGGCGCGTCAGCGAAGCAGC 1105
Db   5474 ACGGCCGGCAAGACGACGCATCGGGCGGCCTCATCGGCCCGGGCTCGAC-CCCAAGGTCACCGGC 5532

Qy   1106 TCGAGACGGCGCCCCATCGTACACCGTCGGCGGGCCCTTCCTCCATTCTAGGCG 1165
Db   5533 CTGGGCAGGCGCCCACGTCGTGTCCCGGACTCGGCGCGCCCACTCGACCATCAAGGCC 5592

Qy   1166 AGCAGTGCCTCACGCCGCGACGGCCGCTCCCGGCGTGGAGGGTGTCTTCAACGAGCCCC 1225
Db   5593 CGCGCGGGGTTGCGGGTTGCGACGTACGACGGGGGTGAGGGAGACCTTCGGGACGCAG 5652

Qy   1226 CTGGTACCCTTAACCGCACATTGACGAGTCTTCTTCACCAAGAACGGGACAGCACC 1285
Db   5653 ATCCCGGCGGGGGGAACCTCAGC------------CCGCGGTTCAACC 5686

Qy   1286 TGGTTGGACTACTTACCACCCCAAGGCGCCAGCAGACACCTGGTTACGGCCGGCACATGGAGGGCACGT 1345
Db   5687 AGGGCCACCAGCTCGAGCCCGGGCAAGGCGCTGTTGTACGACGGGCAGCCTGACCGTGC 5746

Qy   1346 ACACCGCGACGAGGAGGACTGCACCTACGGAGTCGGCGCCTCGTCGTCTCGGGACAGGGCAAAGG 1405
Db   5747 CCGCCGACGGGGGAGTAGGCCATCGCGGGGTCCGTGCGGGTGGTTACGCCACGGTGCAG- 5805

Qy   1406 CGTACGTAGACGACGAGCAGCTCGTCGTCGACAAACGGTCCCCGGCGATGGCT 1465
Db   5806 -------------------------------CTCGGCAGCCACA 5818

Qy   1466 TCTTCGGCGCTCCGCGCACCCGGAGGAGGACGGGGCCGATCAAATCTCGTTCAAGGGCAACACGT 1525
Db   5819 CCATCGAGGCGCCGGTCAGGTCTCACGGCAAGGTGAGCACCCCGCTCCTTCAAGCTGACCAAGG 5878

Qy   1526 ACAAGTTCAAGATCAGAGTTCGGCGCTCCGCACCCACCTACACCCTCAAAGGGCGGCGACACCATCG 1585
Db   5879 GCACGCACAAGCT-------------CACGATCTCGGGCTTCGCGATGA 5914

Qy   1586 TCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCCAAGGTCATTGAGGAGCCGCCCGAAA 1645
Db   5915 GTGCCACCCCGGCTCTCCCTGGGAACTGGGCGGTGGACGGGCGCGCGCGGACAGCGA 5974

Qy   1646 TCGGAAAAGTCCGTGCTCGCCCCTCGCGCCAAGAGCACGACAGTCATCACTCTGCGCGGGCCTTA 1705
Db   5975 TCGGGAAGGCCGTGGAGTGCGGCCGCCGGAAGGGCCCGTACGGCGGGTCGGTGCTTGCC----- 6026

Qy   1706 ACGGCGACTGGGAGAGGCCGCCGACCGGGCGGCGATGAAAGCTCCCCGGCGGTGCTGG 1765
Db   6027 -CTRACGACGACGGCACCGAGGGGTCGACCGTGTCCGACAACCGTGTGGTCCTCAACACG 6085

Qy   1766 ACCAGCTCATTGCCGACGTGGGCCCGGCCGCCGACGACAAACCTGTCGTCGCTGCTCATGCAGACGG 1825
Db   6086 ACAAGCTGATCTCGGCTGTGTCGACGCCAACCGGCGACCGCTCGGGCCTGGTACG 6145

Qy   1826 GCACCGCCGAGGAGAGATGCCCTGGCTCGACGCCACGGCCCTGGTACC 1885
Db   6146 GTTCGTCGGTGCTGATGCCGTGGCTGCGGGCTCCGGACTACAACCCCTCGG 6205

Qy   1886 GCCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTTGGCGACTACACACCCTCG 1945
Db   6206 CGGGCCCAGGCGCGGGGCGGCCCGAGCGCCACCGGCCGCTGACGTCAACGCCGACG 6265

Qy   1946 GCAAGCGTGTCCCTCAGCTTCCCC-----------AAGCGCTGCAGGACAACCCGCGTTTC 1996
Db   6266 GCAAGCTCACGCAGCAGAGCTTCCCGGCGCCGGAGACCACGCGGCGGTCCGGCGACCCGA 6325

Qy   1997 TCAACTTCCGCACCGAGCCGCGGCGGCCGCACCGGCTGTACGGCGGGAGGACGTTCTACGTCGGGTACA 2056
Db   6326 CAAGCTACCGGCGGCGTCGACAACCACCAGCAAGACGACGTACCGCGAGGGCATCCACGGTCGGGTACC 6385

Qy   2057 GGTACTACGGAGTTTGCCGACAAGGACGTCAAATTTCCCCTTTGCCACGGCCGTGTCCTACA 2116
Db   6386 GCTGGTTCGACAACGGAGAACGTCAAGCCGCGCTGTTCCCGCGCTCACGGCACAGCCGTGTACA 6445

Qy   2117 CCACTTTTGCCTTTTCCAATCTCTCCGGTGTTCGACAAGGAC---GGCAAGGTGAGCGTGT 2173
Db   6446 CCTCGTTCACGCACGAGCGCCCCCGACACCGTCGTGGCTGCCTACTCCACGGGTGCTGTCA 6505

Qy   2174 CCCTCTCCGTGAAGAACACACCGGTCGCCCGGCGCACAGGTGGGCCCAGCGTCTCACGTCA 2233
Db   6506 CGGGTCACGGTGCGGCAACAGCGGGAAGGCCGCCAGGAGGTGCTGACGGCCGTGTACG 6565

Qy   2234 AGCCCCTCCAAGCGCGGCCAAGAGATTAACCGCCGCCGTCAAGGACGCCTTCGCAAAGG 2293
Db   6566 GTGCCAGACCCGAACCGTGACGGCTCCCAGGCGAAGAGAAGGCTCGTGGGCTACACGAAGG 6625

Qy   2294 TCGAACTGCAGCCCCGGCGACGCGAAGGCGGTGAC 2327
Db   6626 TCTCCGCTCGCGCCGCGGCGGGCGAGGCGAAGACGGTGAC 6659
```

FT   CDS   /*tag= h
FT   /product= "PikB gene cluster protein #8 (AAY77209)"
FT   complement (12342..13799)
FT   /*tag= i
FT   CDS   /product= "PikB gene cluster protein #9 (AAY77210)"
FT   complement (13706..15043)
FT   /*tag= j
FT   /product= "PikB gene cluster protein #10 (AAY77211)"
FT   CDS   15404..15574
FT   /*tag= k
FT   /product= "PikB gene cluster protein #11 (AAY77212)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU) UNIV MINNESOTA. .
XX
PI   Sherman DH,   Liu H,   Xue Y,   Zhao L;
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 32; 438pp; English.
XX
XX

CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   used to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents, ,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g. fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the desosamine
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC   15439, as given in figure 32.
XX
SQ   Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other;

Query Match   6.7%;   Score 167.2;   DB 21;   Length 13613;
Best Local Similarity   50.4%;
Matches 496;   Conservative   0;   Mismatches 473;   Indels 15;   Gaps 3;

Qy   1359 GGACTGACCTACGAGCTCGGCCTCCTCGTCTCGTCGCACACGGCAAAAGGCGTAGCTAGAGAGCGA 1418
       ||| || |||||| | ||||| | || |||||
Db   3516 GGCGTTCAACCAGGGCCACCAGCTCGAGCCGGCGCAAGGCGCGGGGCGCCTGTPACGACGGCAC 3575
Qy   1419 CCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCCCGGGCGGATGCCTTCTTCGGCTCCGC 1478
       ||| || |||| ||| || || ||||| |||| || | ||||| | | |||
Db   3576 GCTGACCGTGCCCGCCGCCGACGGCGACGGAGTACCGGCATCGGGTCCGGTGCCACCGGTTACGGC 3635

Qy   1479 CACCCGCGAGGAGACGACGGCCGGCATCAATCTCGGCAACGGCCAACAGTTCAAGAT 1538
       ||| ||| ||| | ||||| || || |||||||
Db   3636 CACGGTGCAGCTCGGCGACGCCACCACCTCGAGCGCGGTCAGGTCTACGGCGAGCAG 3695

Qy   1539 CGAGTTCGGCGTCCGCGACCCACCTACACCCTCAAGGGCGACACCACGTCCCCGGCCACGG 1598
       ||| ||| || || || | | || || ||||||
Db   3696 CCCGCTCCTCAAGCTGACCAAGGGCACGCCACAAGCTCACGATCTCGGGCTTCGCGATGAG 3755

Qy   1599 CTCCCTCGCCGTCGGCCGGCTGCAAGGTCATTGACGACGGGCGAAAATCACGAAAAGTCCGT 1658
       || |||| | || |||| | ||||| | ||| ||| |||
Db   3756 TGCCACCCCGGCGCTCCCTGGGAGCTGGGGCTGGGGTTGGACGGGCGGGGCGGGACGACGGAT 3815

Qy   1659 CGCCCTCCGCCAAGGAGACACGACCCAGGTCATCATCTCGCGCGGCCCTTAACCG---CGACTG 1715
       ||| ||| || | ||||||
Db   3816 CGCGGAAGGCCGTGGAGTGCGGCGGGAAGGCCCGTAGGGGTCGGGCTGCTTCCGCCTACGACGA 3875

Qy   1716 GGAGACCGAGGGCGCGCGCCGACCCGCGGAGCATGAAAGCTCCCGGGCGGTTGGGACCAGCTCAT 1775
       || |||||| || | || | ||| |||| |||| || |||
Db   3876 CGGCACCGAGGGCGCGTCGACGTCCGAACCCTGTGCGCTGCCCGGTAGGCAGCAGCAAGCTGAT 3935

Qy   1776 TGCCGACCTGGGCCGGCCGCGGAACCCAAACACGGTGCTCGTCGCATGAGACGGGGCACCCCGA 1835
       ||| || || | || |||| | || || ||||| |||
Db   3936 CTCGGCGTGTCGCGGGAGGCCACAACCCGGAACACGATCGTGGTCCTCAACACCGGGTCGTCGGT 3995

Qy   1836 GGAGATGCCCTGGCCTCGACGCCACGCCCCGGCGCGTCATCCAGGCCCTTGGTACGGCGGCAACGA 1895
       ||||| || | | ||| ||||||
Db   3996 GCTGATGCCGTGGCCGTGTCCAAGACACCCCGGGGGTCCTTGGACATGTGGTACCGGCGCCAGGC 4055

Qy   1896 GACGGGCAAACTCCCATTGCCGACGTCGTCTTTGGCGACTACGAACCCTCGGGCAAGCTGTC 1955
       | ||||| |||||||| |||||| ||||||||
Db   4056 GGGCGCCGAGGCCACCCGGCGCTGCTCTACGGTGACGTCAACCCGACGCCAAGCTCAC 4115

Qy   1956 CCTCAGCTTCCCC---------AAGCGCCTGCAGGACAACCCCGGCGTTGTCTCAACTTCCG 2006
       |||||||||| ||| || |||||
Db   4116 GCAGAGCTTCCCGGCGCCCGAGAACCCAGCACGCGGTCGCCGGGGACCCGGACAAGCTACC 4175

Qy   2007 CACCGAGGCCGGCGGGCGCAGCGTCGTACGGCGGAGGACGTCTACGTCGGGTACAGGTPACTACGA 2066
       ||| || |||||||| | || || || |||| |||
Db   4176 GGGCGTCGACAACCAGCACGACGTACCGGCACGGTCGGGTACCGCGGGCACCGGA 4235

Qy   2067 GTTTGCCGACAAGGAGCGTCAAATTTCCCCCTTGGACGACCGGCCGGCGCCTTCCTCACACACCACTTTTGC 2126
       || |||| ||||| || |||||||
Db   4236 CAAGGAGAACGTCAAGCCGCTGTTCCCGTTCGGGCGCACGGCCGGTCGTGCGGTCAC 4295

Qy   2127 CTTTTCCAATCTCTCCGTGTCTCACACAGGAC---GGCAAGCTCGAGCGGTGTCCCTCTCCGT 2183
       |||||| | ||| || || |||
Db   4296 GCAGAGCCCGCACCGGTCGTGGGTACTCCACGGGTGGTCGTCGAAGTGCACGGTCACGGT 4355

Qy   2184 GAAGAACACCGGCGTCCGTGCCCGGCGCACAGGTGGCCCGGCACGGCTCTACGGTCAAGCGCCCTCCA 2243
       ||||| || | || |||||
Db   4356 CCGCAACACGGGGAAGGCGCCCGGCGCCAGGAGTGTCGCCAGGGCTGCTGGTGTCGGCCAGCCC 4415

Qy   2244 AGCGGCCAAGGATTAACCGCCCCCGTCAAGGAGCGTTCGCAAAAGGTCGAACTGCA 2303
       ||| || ||||| |||||| || |||
Db   4416 GAACGTGACGGCTCCGCAGGCGGAAGGAGAAGAGCTCGTGGGCTACACGGACGAAGGTCGGCTGCC 4475

Qy   2304 GCCCGCGGACGACGAAGGGCCGTGAC 2327
       | |||||| |||||| ||||||
Db   4476 CGCGGGCGGAGGCGAAGACGGTGAC 4499

RESULT 11
AAT32999
ID   AAT32999 standard; DNA; 2256 BP.
XX
AC   AAT32999;
XX
DT   18-OCT-1996 (first entry)
XX
DE   Chimaeric thermostable beta-glucosidase coding sequence.
XX

```
KW   Chimaera; beta-glucosidase; Cellvibrio gilvus; Agrobacterium tumefaciens;
KW   homology; decomposed cellulose; glucose; cellobiose; cellotriose;
KW   cellotetraose; cellopentaose; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..2076
FT                   /*tag= a
FT                   /note= "sequence derived from Cellvibrio gilvus
FT                   beta-glucosidase gene"
FT   misc_feature    2077..2256
FT                   /*tag= b
FT                   /note= "sequence derived from Agrobacterium tumefaciens
FT                   beta-glucosidase gene, encodes amino acids
FT                   759-818 of A.tumefaciens beta-glucosidase"
XX
PN   JP08131168-A.
XX
PD   28-MAY-1996.
XX
PF   09-NOV-1994;   94JP-0299049.
XX
PR   09-NOV-1994;   94JP-0299049.
XX
PA   (NORQ ) NORINSUISANSHO SHOKUHIN SOGO.
XX
DR   WPI; 1996-311636/32.
DR   P-PSDB; AAR97199.
XX
PT   Thermostable chimaeric beta-glucosidase - useful for prodn. of
PT   cellobiose from decomposed cellulose
XX
PS   Claim 2; Page 5-8; 9pp; Japanese.
XX
CC   This is the nucleotide sequence encoding a chimaeric beta-glucosidase
CC   enzyme composed of amino acids 1-692 of Cellvibrio gilvus
CC   beta-glucosidase and amino acids 759-818 from the Agrobacterium
CC   tumefaciens beta-glucosidase, replacing amino acids 693-752 of the
CC   Cellvibrio sequence. The chimaera was constructed by isolating the
CC   C.gilvus beta-glucosidase gene and comparing its sequence with the
CC   corresp. genes from several other species, e.g. Hansenula anomala,
CC   Ruminococcus albus, Butyrivibrio fibrisovens, Agrobacterium tumefaciens,
CC   etc. The homologous sequences from the other species could then be used
CC   to generate chimaeric glucosidase genes. The novel chimaeric protein
CC   acts on decomposed cellulose to generate glucose and cellobiose and has
CC   a reduced ability to decompose cellobiose as compared to cellotriose,
CC   cellotetraose or cellopentaose as substrate. The novel enzyme has an
CC   optimum pH of 6.0 and temp of 40-45 deg. C, is stable in pH 4-9 and at
CC   35 deg. C.
XX
SQ   Sequence 2256 BP; 370 A; 786 C; 763 G; 337 T; 0 other;
```

```
Query Match     6.4%;   Score 161;   DB 17;   Length 2256;
Best Local Similarity   52.4%;                    Pred. No. 1.1e-20;
Matches  449;   Conservative  0;   Mismatches 375;   Indels 33;   Gaps 3;
```

```
Qy   1598 GCTCCCTCCGGGTTCGGCGGCGGCTGCAAGGTCATCAGGCCTTAACCGCGACTGGG 1657
                                                                            
Db   1346 GCGCGATTCAGGCCCCAGGCCCCGAAATGCGAAAGTCGTTCATTGAGCCGGACGGCCAG 1405

Qy   1658 TCGCCCCTCCCCAAGGAGCACGACCAGGTCATCACTGCGGCCGCCCTTAACGCTGGG 1717
                                                                            
Db   1406 CCCCGCGCCGGCGCCGGTGGCCGCCGCGCACTCGCGCTTCGCCACCAGTGGA 1465

Qy   1718 AGACCGAGGGGCGGCCGCCGCACCCGCGGGCGAGCATGAAGCTCCCGCGGGGTGCTGACCAACTCATTG 1777
                                                                            
Db   1466 TCGGCGGAGGCCCAAGCGACACGGCGCTCGCGCTTCGCCGCCAACCAGTGGA 1525

Qy   1778 CCGACGTGGCCGCGCCGGCGAACCCAAACACCGTCGTGCTTCGTTCGTCATTGCAGACGACGCCGTCGATCA 1837
                                                                            
Db   1526 CGTCGGTGCGCGGCCGCGGCGCCAACGGCGCGCCACCGGTCGTGGTTCGTCGTGCTGCAGAACCGGCACGGGTCA 1585

Qy   1838 AGATGCCCTGGCTCGAGCGCCAGCACCCCGCCGTCATCCAGGCCTGGTACGGACGGCGCAACGAGA 1897
                                                                            
Db   1586 CCATGCCCTGGCTGGCGGCGCCGGTTCCGGCGGCGTGCTGGGAAGCCTGGGTATCCGGGCCACCAGCG 1645

Qy   1898 CGGGCAACTCCATTGCCGACGTGCGTCTTTGGGGACTTACAAACCCCTCGGGCAAGCTGTGCC 1957
                                                                            
Db   1646 GCGGCGAGGCGATCGGCGAAATGTGCGTGTTTTGGGCGGTCAAATCCGTCGGGCCACCTGCCGG 1705

Qy   1958 TCAGCTTCCCAAGCCGCCTGCAGGACAACCCCGGTTTCTCAACTTCCGCACCACCACTT 2017
                                                                            
Db   1706 CCACCTTCCCGCAATCCGAGCAGCAACTGGCGGCGCCGAAACTTCGATGGCGATCCGAGA 1765

Qy   2018 GGGGCACCGCTGTACGGCGGAGGACGGTCTAC-------------GTCGGGTACAGGTTACT 2062
                                                                            
Db   1766 ACCCCGAGCGTGCAGTTCGGCCGTCGACCTACCACGAAGGTGCGGCGGCGGGGTCGGGCTACACAAGTGGT 1825

Qy   2063 ACGAGTTTGCCGACAAGGACGTCAAATTTCCCCGTTGGCCACGGCGTGCCTACAACCACTT 2122
                                                                            
Db   1826 TCGACCTGAAGGGCCACAAGACCCGCGCTTCCGGTTCCGGCGACGGCGGCTGTTCCCTACAACACCT 1885

Qy   2123 TTGCCTTTTCCAAATCTCTCCGTGTGTCTCACAAGGAGCACGGGCAAGCGTGAGCGTGTTCCTCTCCG 2182
                                                                            
Db   1886 TCGCGTACTCCGGTTGTTGTCCGGCCAGCTCAAAGGATGGCCCGCCTGCACGGTGGCGCTTCAAGG 1945

Qy   2183 TGAAGAAACACCGGGCTCCGTGCCCGGCGGCACAGGTGGCCGCACGCTCTACGGCGCCCTCC 2242
                                                                            
Db   1946 TGACCAAACACCGGCAAACGTGGCCCGGCCAAGGCACGGTGCCCGGCGTGACCGGCCGGAT-- 2003

Qy   2243 AAGCCGCCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGGTCGGAACTGC 2302
                                                                            
Db   2004 ----GTCCACCAAATGGGAGGGCGCCTGGCCGGAAGGGCGGCGGCTTCAAGG 2059

Qy   2303 AGCCCGGCGAGACGAAAGGCGCCGGTGACAAATCGAGGAGGCGAGAGAGTACGTCGCT---- 2355
                                                                            
Db   2060 TGCCCGGCGGAGACCAAGCGCCGGGCGCCGACCGGTACGGCGGCCGCC 2119

Qy   2356 ----GGGTATTTTGATGAGGAGGACGGATCAGTGGTGTGTCGGAAAGGGGTGCACTATTGAGG 2410
                                                                            
Db   2120 ACTTGGCTTACTTCGGATGTCGAGGCCGGGCGGTCGTTCCGGGTGATGCGGGCGCAAGTACGAGC 2179

Qy   2411 TTATCGGTGAGCGGACAGC 2427
                                                                            
Db   2180 TGATCGGTGGCGGCCAGC 2196
```

```
RESULT 12
AAA09469/c
ID   AAA09469 standard; DNA; 50937 BP.
XX
AC   AAA09469;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Streptococcus oleandomycin gene cluster.
XX
KW   oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW   PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW   acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW   macrolactone; antibiotic; motilide; erythromycin; ss.
XX
OS   Streptococcus antibioticus.
XX
FH   Key             Location/Qualifiers
FT   CDS             152..1426
FT                   /*tag= a
FT   CDS             complement (1528..2637)
FT                   /label= oleI
FT                   /*tag= b
FT                   /label= oleN2
FT   CDS             complement (2658..4967)
FT                   /*tag= c
FT                   /label= oleR
FT   CDS             5772..18224
```

```
FT              /tag= d
FT    CDS       /label= ORF1
FT              /product= 8,8a-deoxyoleandolide_synthase_1
FT              /tag= e
FT    18267..29717
FT    CDS       /label= ORF2
FT              /product= ORF2
FT    29787..40346
FT              /tag= f
FT    CDS       /label= ORF3
FT              /product= 8,8a-deoxyoleandolide_synthase_2
FT    40625..41830
FT              /tag= g
FT    CDS       /label= oleP1
FT              /product= 8,8a-deoxyoleandolide_synthase_3
FT    41878..43158
FT              /tag= h
FT    CDS       /label= oleG1
FT    43163..44443
FT              /tag= i
FT    CDS       /label= oleG2
FT    44433..45173
FT              /tag= j
FT    CDS       /label= oleM1
FT    45251..46411
FT              /tag= k
FT    CDS       /label= oleY
FT    46491..47714
FT              /tag= l
FT    CDS       /label= oleP
FT    complement (47808..49517)
FT              /tag= m
FT              /label= oleB
XX
SQ    Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;
```

CC   modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC   promoter. Also discussed are recombinant oleandolide PKS in which the
CC   module 1 KS domain is inactivated by deletion or other mutation. In
CC   particular, the inactivation is mediated by a change in the KS domain
CC   that renders it incapable of binding substrate (the KS1-o mutation),
CC   rendered by mutation in the codon for the active site cysteine. The
CC   oleandolide PKS is useful for synthesizing polyketides, which are useful
CC   as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC   in host cells such as Streptomyces coelicolor and S. lividans is also
CC   made possible. Unmodified oleandolide compounds can be provided to
CC   cultures of Saccharopolyspora erythraea and converted to the
CC   corresponding derivatives of erythronycins A-D.

CC   This is part of the Streptococcus antibioticus oleandomycin gene cluster.
CC   The oleandolide polyketide synthase (PKS), also known as
CC   8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC   (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC   "modular" enzyme, where each ORF encodes 2 extender modules and
CC   the first ORF also encodes the loading module. Each module is composed
CC   of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC   acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC   contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC   glutamine, present instead of the active site cysteine required for
CC   activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC   of polyketide macrolactones through multistep pathways involving
CC   decarboxylative condensations between acylthioesters followed by cycles
CC   of varying beta-carbon processing activities. The macrolide product of
CC   the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC   glycosylation to yield oleandomycin, an antibacterial polyketide. The
CC   invention concerns an isolated recombinant DNA compound, comprising a
CC   coding sequence for a domain of loading module or any one of extender

CC   Recombinant DNA compound encoding oleandolide polyketide synthase for
CC   synthesizing polyketides comprising a coding sequence for a domain of a
CC   loading module or any one of extender modules

Query Match   5.5%;   Score 138.6; DB 21; Length 50937;
Best Local Similarity 54.0%;   Pred. No. 1.8e-16;
Matches   354; Conservative   0; Mismatches 289; Indels 12; Gaps 3;

```
Qy   1710 CGACTGGGAGACCGAGGGCGGCCGAACCGCCGAGCATGAACCTCCCCGGCGTGCTGGACCA 1769
           |||| |||| ||||| || |||| | | |||||| |||||| || |||||| || ||||||
Db   3417 CGACGACGGGCTCGGAGGACGGGGGACCCCACGTCGCTGGCTGTCCCCGGCCCAGGACGA 3358

Qy   1770 GCTCATTGCCCGACGTGGCCCGCGCGAACCCAAACACCGTCGTCGGGTACAGGGGCAC 1829
           || || |||||||| || |||||||||||| ||||||||||| |||||| |||||||
Db   3357 CCTGATCAGCCGCGCCGTGGCCCGGGGGTCAAACCACGACCGTGGTGTGCTCAAACACGGGCTC 3298

Qy   1830 CCCCGGAGGAGATGCCCTGGCTCGACGCCACGCCCGCGCATCCCAGGCGGTACGGCGG 1889
           |||| ||||||| ||| |||||||||||| || ||||| ||||||||||||||||||
Db   3297 CTCCCTCACCATGCCCTGGGCTGAGGAAGAACCGCCCGGGTGCTCACCAATGTGGTACCCGGG 3238

Qy   1890 CAACGAGACGGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCTCGGGGAA 1949
           |||||||| || || |||||||||||| ||||| || |||||||||||||||| |||||
Db   3237 CCAGGCCGGCCCCGAGGGGACCACCGCCCCTCGCTGTTCCGGCGACCCGGGCCGACCCAGGAGCC 3178

Qy   1950 GCTGTGTCCCTCAGCTTCCC----------CAAGCGCCTGCAGGACAACCCCGCGTTTCTCAACTT 2003
           |||| ||||||||||||||          |||||||||||||||| |||||||| ||||||||||
Db   3177 GCTGACCCAGACTCTTCCCGGCCCGGGGGACGGAGGGGACGACGCCGGTCCGCCGGGGGGACCGCAGCG 3118

Qy   2004 CCGCACCGAGGCCGCGGCGCACGCTG---TACGGGCGGAGGACGTTCTTACGTCGGGTACAGGTA 2060
           || || |||||||||||| |||||||   ||||||||||||||||| || ||||||||||||||
Db   2997 CTTCGACTACCGGGACCTGAAAGTTGACGGCGGACGGGGACCTTCGACGTCTGCGGTTGCTT 2938

Qy   2061 CTACGAGTTTGCCCGACAAGGACGTCAAATTTCCCCTTTTGGCCCACGAGCCGTCCTTACACCAC 2120
           | |||||||||||||||||| || ||||||||||||||| || ||||||||||| |||||||
Db   2937 CACGCTGGGCCAACACCGGTACGGCACGCGCAAGGAGGTCCCGCCAGGGTCTACGTGGGGCC 2878

Qy   2121 TTTTGCCTTTTCCAATCTCTCCGGTGTCTCACAA---GGACGGGCAAGCTGAGGGTGTCCCT 2177
           |||| |||| ||||||||||||||||||| |||   ||| ||||| ||||||||||||||||
Db   2877 GTCCCGCACGTGCGGGTTGCGGCAAGCGTGCCCTGGCGGGGCCTTACGGGAAGGTCGA 2818

Qy   2178 CTCCGTGAAGAACAACCGGCGCTCCGGTGCCGGGCGCCAGCTGGCCGCTCTACGTCAAGCC 2237
           |||| |||||||||||||||||||||||||||||||||||||||||| |||| |||||||
Db   2817 GCTGCGGCGCGGGGCGAGCAGCAGGAGGCTGACCCTGCACGTCGCAGGCGCCCT 2763
```

XX
KW   Glycosidase; thermostable; textile; food processing; pharmaceutical;
KW   detergent; baking; industry; Thermococcus; Staphylothermus;
KW   Pyrococcus; glucose; soluble oligosaccharide; ss.
XX
OS   Thermotoga maritima.
XX
FH
FT   Key   Location/Qualifiers
FT
FT   CDS   1..2166
FT   /*tag= a
FT   /product= Glycosidase
XX
PN   WO9725417-A1.
XX
PD   17-JUL-1997.
XX
PF   10-JAN-1997;   97WO-US00092.
XX
PR   13-SEP-1996;   96US-0712612.
PR   11-JAN-1996;   96US-0583787.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Bylina EJ,   Lam DE,   Mathur EJ,   Swanson RV;
XX
DR   WPI; 1997-372858/34.
DR   P-PSDB; AAW34558.
XX
XX
PS   Claim 4; Fig 5; 82pp; English.
XX
CC   The present sequence encodes glycosidase isolated from Thermotoga
CC   maritima. The enzyme or its encoding nucleic acid sequence is used for
CC   generating glucose from soluble oligosaccharides. The enzyme can be
CC   used in the food processing, pharmaceutical, textile, detergent and
CC   baking industries. The enzyme is also used to treat lactose intolerance,
CC   as a diagnostic reporter molecule, in corn wet milling or in the fruit
CC   juice industry. The enzymes can be used to hydrolyse guar gum to remove
CC   non-reducing terminal mannose residues. The nucleic acids encoding the
CC   enzyme may be used to generate probes to identify similar sequences.
XX
SQ   Sequence 2166 BP; 668 A; 454 C; 578 G; 466 T; 0 other;

Query Match   5.4%;
Best Local Similarity   53.5%;
Matches   284;   Conservative   0;   Mismatches   247;   Indels   0;   Gaps   0;

Score 135.8;   DB 18;   Length 2166;
Pred. No. 4.9e-16;

```
Qy      207 GCTCGGTTCCACATTCAACCAAAACTCTGCTCGAAGAGCGCAGGTAAGATGATGGGGCAAAGA 266
                | ||||||| || | ||| ||  |  || |||  | ||||||| | | ||||||| |||||||
Db      258 GCTCGGCTTCTACCTGGAACAGAGACCTTCTGGAAGAAGTGGGAAAAGCCATGGGAGAAGA 317

Qy      267 GGCCATCCCTAAGAGAGTGGGCCATGTGATCCTCCGGCCCCGACTATCAACAATGCAACGCTCCCC 326
                |  |||   ||||||| ||| ||||| ||||||||| |||||| |||||||||||| |||
Db      318 AGTTAGGGAATACGGTGTCGATGTGCTTCTTGCACCTGCGATGACAATTCACAGAAACCC 377

Qy      327 TCTCGGGTGGACGTGGCTTCGGAGTCGATTGGTGAGGATCCGTTTCCTGGCGGGCTTGGGAGC 386
                ||| ||||| |||| |||| |||| || ||| || || | || |||||| |||||| |||
Db      378 TCTTTGTGGAAGGAAATTTCGAGTACTACTCAGAGAAGATCCTGTGTCCTTTCCGGTGAAATGGC 437

Qy      387 TGCGGCCTCTCCATCCCGCGGCCATTCAGACACTGGAGCTACGATCAACGACACTTTTT 446
                || ||||| ||  || | || | ||||| || || || ||||| || ||||| ||||||
Db      438 TTCAGCCTTTGTCAAGGGAGTTCAATCTCAAGGGGTGGGAGCCTGCATAAAACACTTTGT 497

Qy      447 GTGCAATGATCAGGAGGAAGGACGACAGGCGCCATGAGTGAGTGCAGAGAGCCATCCGTCACGGAGCGGGCTCT 506
                ||| || || || ||||| || || ||||| || ||||| | ||||| |||| |||||| ||| |||||||
Db      498 CCGGAACAACCAGGAAACGAACACAGGATGGTAGTGGACACGATCATCGTGTCCGAGGCGAGCCCT 557

Qy      507 CCGTGAAATCTACCGCACTCCCGGTTCCCAGATTGCTGCTGCGAGAGGAGGTGCGGTT 566
                | ||||||||| |||| |||||||||||| ||| ||||||| ||||||||||||||||
Db      558 CAGAGAAATATATTACCGACTCCTGGTTTGAAAATTGCTGTCAAGAAAAGCAAGACCCTTGGACCGT 617
```

Db   558 CAGAGAAATATATTACCGACTCCTGGTTTGAAAATTGCTGTCAAGAAAAGCAAGACCCTTGGACCGT 617

RESULT 14
AAV36911
ID   AAV36911 standard; DNA; 2166 BP.
XX
AC   AAV36911;
XX
DT   21-DEC-1998   (first entry)
XX
DE   Thermotoga maritima MSB8-6G glycosidase gene coding region.
XX
KW   Glycosidase; MSB8-6G; thermostable enzyme; oligosaccharide;
KW   glucose; sugar; baking; textile; detergent; beta-galactosidase; ss.
XX
OS   Thermotoga maritima strain MSB8-6G.
XX
XX
PN   WO9824799-A1.
XX
PD   11-JUN-1998.
XX
PF   08-DEC-1997;   97WO-US222623.
XX
PR   10-OCT-1997;   97US-0949026.
PR   06-DEC-1996;   96US-0056916.
XX
PA   (DIVE-) DIVERSA CORP.
XX
PI   Bylina EJ,   Lam DE,   Mathur EJ,   Swanson RV;
XX
DR   WPI; 1998-362407/31.
DR   P-PSDB; AAW49862.
XX
XX
PS   Claim 1; Fig 5a-b; 92pp; English.
XX
CC   This isolated polynucleotide comprises a coding region for
CC   glycosidase MSB6-6G (see AAW49862) from a Thermotoga maritima MSB8
CC   clone (6G) that grows optimally at 85 degC in high salt medium.
CC   The sequence shows 53% nucleic acid identity to beta-galactosidase
CC   B of Clostridium thermocellum. The invention provides 18
CC   polynucleotides (see AAV36907-24) coding for thermostable glycosidases
CC   (see AAW49858-75) having glucosidase, alpha-galactosidase,
CC   beta-galactosidase, beta-mannosidase, endoglucanase
CC   or pullulanase activity. Vectors and host cells are also claimed.
CC   A method is provided for producing the enzymes by recombinant
CC   techniques. A claimed method for generating glucose from soluble
CC   cell oligosaccharides comprises contacting a sample (selected from
CC   dairy products, fruit juice, textile, detergent, textile, guar gum, animal
CC   feed, plant biomass or waste product) containing oligosaccharides
CC   (selected from maltose, cellobiose, lactose, sucrose, raffinose,
CC   stachyose, verbascose, cellulose, starch, amylose, glycogen,
CC   disaccharides, polysaccharides and pullulan) with one of the
CC   claimed glycosidases such that glucose is produced.
XX
SQ   Sequence 2166 BP; 675 A; 444 C; 581 G; 466 T; 0 other;

```
Qy      567 CATGACGGCGGTACAATGGCATCAATGGCGTGCGTGTCGTGTTGCACGGAGAGAACCCTAAATATCTTGA 626
                |||| || |||||   || ||||| ||||||| |||||| || | |    ||||||
Db      618 GATGACGGCCGTTACAACAAACAAACTGAATGGAAAATACTGTTCACACAAGAAGCTTTGAA 677

Qy      627 TGGGATGGCTTCGAAAAGGAAGGGATGGGGGGTTGGGGATGGCCATAATCATGAGGCACTGTGTACGGCAC 686
                ||||| | || || ||| ||||| ||||| |||| ||||||||||| | ||||||||| |||||||
Db      678 GAAGGTTCTCAGGGAAGGAAAATGGGGGATTTTGGCGGTGGTTTTCGTGATGAGGCGACTGGGTTACGCGGGG 737

Qy      687 ATACAGTACCACACAGAAGCCGTTGTGGCAGGCCTCGAACCTCGAGAGTGCCCGG 737
                |||   ||||| |||||| ||||   |||||| |||||||| |||||||||| |||
Db      738 AGACAACCCTGTTAGAACAGCTCAAGGCCGGAAACAGATATGATCATGCCTGG 788
```

Db   738 AGACAACCCTGTTAGAACAGCTCAAGGCCGGAAACAGATATGATCATGCCTGG 788

RESULT 14   (continued)

Qy   Glycosidase enzymes from organisms of the genera Staphlothermus,
PT   Pyrococcus and Thermococcus - for deriving sugar from
PT   oligosaccharides, useful in the e.g. food processing, textile or
PT   baking industries

```
Query Match      5.4%;    Score 135.8;  DB 19;  Length 2166;
Best Local Similarity   53.5%;  Pred. No. 4.9e-16;
Matches 284;  Conservative  0;  Mismatches 247;  Indels  0;  Gaps  0;

Qy    207 GCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAGA 266
          |||||.|| | || .|| || | ||||||| ||| ||||||||||.|||||||||||||
Db    258 GCTCGGCTTCTACCTGGAACAGAGACCTTCTGGAAGAAGTGGGAAAAGCCATGGGAGAGA 317

Qy    267 GGCCATCGCTAAGAGTGCGCATGTGATCCTCCGCCCGACTATCACACATGCAACGCTCCCC 326
          | || || |  ||  ||||||||||||||||||||||||||||||||||||||||||||
Db    318 AGTTAGGAGAATACGGTGTCGATGTGCTTCTTCGCACCTGCGATGAAACATTCACAGAAACCC 377

Qy    327 TCTCGGTGGACGGGACGTGGCTTCGAGTCGATTGGTGAGGATCGTTCCTGCGCGGGCTTGGGAGC 386
          ||| |||||| || ||| ||||||||||||||||||||||||||||||||||||||||| ||
Db    378 TCTTTGTGGAAGGAAATTTCGAGTACTACTCAGAAGATCCGTCCTTTCCGGTGAAATGGC 437

Qy    387 TGCGGCGTCTCATCCGCGGCGCATTCAGCAGCCACTGGAGTTCAGGCTACCATTCAAGCACTTTTT 446
          | || ||||||||||||||||||||||||||||||||||||||| ||| ||| ||| |
Db    438 TTCAGCCTTTGTCAAGGGAGTTCAAATCTCAAGGGGTGGGAGCCTGCATAAAACACTTTGT 497

Qy    447 GTGCAATGATCAGGAGGACAGCGCGCCATGATGGTGCAGAGCACGCGGAGGGGCTCT 506
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    498 CCCGGAACAACCAGGAAACGAACAGGATGGTAGTGGAGACGATCGTGTCCGAGCGAGCCCT 557

Qy    507 CCGTGAAAATCTACGCACTCCCGTTCCAGATTGCTGTGGCGGAGGAGACTCCCAGCCGGGGTGCGTT 566
          | ||||||||| || |||||||||||||||||||||||||||||||||||| || ||
Db    558 CAGAGAAAATAATATCGTAGAAAGGTTTTGAAAATTGCTCTGTCACAAAGAACAGACCCTGGACCGT 617

Qy    567 CATGACAGGCGGTACAAATGGCATCAAATGGCGTGTGCGTGCGCAGCGAGAACCCTAAATATCTTGA 626
          |||||| |||| || |||||| |||||||||||||||||||||||||||||||||||||||||
Db    618 GATGAGCGCCTTACAACAAAACTGAATGGAAAAATACTTCTTCACAGAACGAATGGCTTTTGAA 677

Qy    627 TGGGATGCTTCGAAAGGAATGGGGTTGGGGATGGCCTAAATCATGAGCGACTGGGTACGGGAC 686
          ||| |||||| ||| |||||||||||||||||| |||| ||||||||||||||||||||||||
Db    678 GAAGGTTTCTCAGGGAGGAAGAAATGGGGATTTGGCGGTTTCGTTGATGACCGACTGGGTACGGGGG 737

Qy    687 ATACAGTACCACAGAGAGCCGGTTGTGGCAGGCCTCGACCTGCGAGGATGCCCCGG 737
          ||| ||||||||||||||||||||||| ||||||||||||||| | ||
Db    738 AGACAACCCTGTAGAACAGCTCAAGGCCCGGAAAACGAATATGATCATTGCCTGG 788
```

```
Query Match      4.5%;    Score 111.6;  DB 22;  Length 4403765;
Best Local Similarity   52.0%;  Pred. No. 2.2e-11;
Matches 345;  Conservative  0;  Mismatches 289;  Indels  30;  Gaps  3;

Qy   1723 GAGGGGCGCCGACCCGCGGGGAGGCAAAGCTCCCCGCGGGGTTGCTTGGGACCAGCTCAATTGCCGAC 1782
          |||||   |  || || ||| |||||||||||||||||| ||  |  || ||||| ||
Db 217690 GAGGGCCTTCGACACGCCGATCTGTCCGCCTGCCATGGGGTCAGGAATGCGCTGCATGCCCGCA 217749

Qy   1783 GTGGCCGCCGCGGAAACCCAAAACACCGTCGTCGTCATCGAGACGGGGGCCCCCGAGGAGATG 1842
          || || |||| |  |||| ||||| ||||| || || ||||||||||| || |||||||||
Db 217750 GTCGCGTCCGCCAACGGGAAATACCGTTGTGGTGCTTGAGAGCGGGGCAAACCCGGTGACCATG 217809

Qy   1843 CCCTGGCTCGAGCCCACCCCCGCCCGTGCATCCAGGCCTGGGTACGCGGCGCAACGAGACGGGC 1902
          ||||||| ||| || ||   || |||| |||||||||||||||||| || |||||||||||||
Db 217810 CCCTGGCCCGACTCGGTTGAACGGCCATCATGCAGGCCTGGTATCCCGGCGGGGGGGTGGC 217869

Qy   1903 AACTCCATTGCCGACGTCGTCGTTTTGCCGACTACTACAACCCTCGGGGCAAAGCTGTTCCCTCAGC 1962
          |||||||||||||| || || || ||||||||||||||||| ||||||| |||||||||||||||||
Db 217870 CAGGCCGTTGCGGAGATTGTGGACCGGCAGGTGAATCCTTCGGGGGCGGGCTGCCGGATCACC 217929

Qy   1963 TTCCCCAAGGCGCCTGCAGGACAAACCCCGCGTTTCTCAACTTCCGCACCGAGGGCCGGGCG 2021
          ||||| || || |||||||||||  ||||||| || || ||||||||||||||||| |||
Db 217930 TTCCCGGTCGGATCTCGGTCAGACGCCACGCTCGCAAACCGCCCGAGCTCGGTGCCCGTGG 217989

Qy   2022 -------------CACGCGTACGGCGAGGAGGACGTTCACGTCGGGGGTACAGGTACTACGAG 2067
                       |||| |||||||||| ||||||| || |||||||||| |  ||
Db 217990 GGGACAATCGACCACGATCCACTACACGAGGGCGCCGAGTTGGTTACCGGCTGGGTTTGCC 218049

Qy   2068 TTTGCCGACAAGGACGTCAAATTCCCCTTTGGCCACGGCGCCTGCCTACACCACTTTGCC 2127
          ||| ||||| ||||| || | ||||| || |||||||||| || || ||||||||||||
Db 218050 AGCACAAATACAGACACCCGATGTTCGGCGTTCTCGGTTCACGGCTGTGCCTATACCAGTTCGAG 218109

Qy   2128 TTTTCCAAATCTCTCCGTGTGTCTCACAAAGGACAGGCAAAGCTGAGCGTTGCCCTCTCCGGTGAAG 2187
          |||||| |||||| ||| || ||||||||| ||||||   ||||||||| || |||| |||||||||
Db 218110 TATCGTGACCTGGTTGGGTGTTGACGGGGC----GGCCACACCGTGCACCGCCAGTTTCAGCGGTTACC 218166

Qy   2188 AACACCGGCTTCCGTCGCCTGCCACAGTTGGCCCCAGCGTCTACGTCAAGCCCTCCAAGCG 2247
          ||||| ||| || || || || ||||||||||||||| |||||||| ||| |||||| |
Db 218167 AACACGGCGGCCAGACCGCAGCGGGCGGCGGATGTCCCGCAGCTGTATATATGACGGCAGCTCCCGGC 218226

Qy   2248 GCCAAGATTAACCGCGCCCGTCAAGGAGCTTCAGGAGGTTCGCAAAGGTTCGAACTGCAGCCC 2307
          || || ||| || || || ||||||||| ||||||| |||| ||| || |||||| ||||||
Db 218227 GAATCGCGGTTGCG--------------GTTGCTGGGGATTGCAGCGTTCGACGGGGTTCGAGCGCTCGAACCC 218274

Qy   2308 GGGCGAGACGAAGGCGGTGACAATCGAGGAGCAGGAGAAGTACGTCGCTGCGTATTTTGAT 2367
          ||| |||||| |||| || ||||||||||||| |||| || |||||||| || || |
Db 218275 GGCCAGACGTCGGCGCGGGTAAGGATCGAGGCGACACCGGCACCCGACGACTCGGCTACGACGGC 218334

Qy   2368 GAGG 2371
          ||||
Db 218335 GAGG 218338
```

us-10-027-000-3.rng